

Europe Region Special Section

DOI:10.1145/3514188

Welcome!

WELCOME TO THE second *Communications* Regional Special Section spotlighting European countries and Israel. On a relatively small portion of the Earth, this region includes almost 50 countries with enormous cultural and socioeconomic diversity that is also reflected in the richness of its business structures and computer science research. The first Hot Topic article in this section illustrates the high overall share of European public research on a global scale, and further highlights significant differences within the region. We are happy to report the authors in this special section represent 15 countries throughout Europe plus Israel.

An important goal emphasized by the European Union (E.U.) and many individual countries is to attain digital sovereignty of the private and public sectors, while further developing areas of traditional industrial and design strengths into the future. Data strategies and regulations by the E.U. therefore emphasize resilient networking of decentralized digital infrastructures in addition to the presence of international big players from, for example, North America or China. In this regard, important initiatives spotlighted here include the GAIA-X initiative on decentralized data space infrastructures, the FENIX network of scientific high-performance computing, the transition from 5G to 6G networks, and more domain-specific initiatives related to Industry 4.0, to robotics, and to energy informatics in the context of de-carbonization—an important element of making the E.U.’s “Green Deal” become a reality.

Europe has shown world leadership in placing people at the core of the digital and AI revolutions, as exemplified by the General Data Protection Regulation (GDPR) and the recently proposed AI regulation where human-centric aspects are at the core, such as bias avoidance, transparency, veracity, and the preservation of human autonomy. This “European approach” in existing or proposed regulations and research funding schemes have significant implications for systems design and computer science research topics. Thus, this section includes several contributions in this regard, such as studies of privacy-preserving networks, fair recommender systems, and general design implications to build trusted AI.

Beyond these strategically driven topics, the section also highlights selected fundamental research and industrial innovation projects in areas of traditional European CS strengths, such as formal methods, process management, and socio-technical user interface design. To bring the emerging field of quantum computing into more complex real-world solutions, its embedding into formally based software engineering methodologies requires deeper study. The generalization of the success of deep learning algorithms to “broad AI” is summarized by one of the top European contributors to this success. Further along on the evolutionary curve,


ILLUSTRATION BY SPOOKY POOKA AT DEBUT ART. FOR CREDITS ON IMAGES IN COLLAGE, SEE P. 3.



the traditional European strength in business process management, exemplified by the early success of SAP, is reaching the next stage by highly successful European-initiated start-ups in process mining and robotic process automation that have their basis in fundamental Europe-driven research. Finally, Europe has a long tradition in human-centric systems and user interface design under special consideration of user diversity and inclusion. Recent developments are highlighted in application-oriented articles for crisis management and chronic patient support, and in a contribution on completely new interface types, such as on-skin computing.

To create this section, we conducted a (virtual) two-day workshop from August 25–26, 2021, with 34 presentations. Participants were in part directly invited, in part preselected from over 100 submissions in response to a public call for abstracts. The lively and fruitful discussions led to the sharpening of our foreseen structures, including a couple of newly formed author teams. In further steps, we converged to a choice of 22 initial articles invitations. After further discussions and a review of submissions by the guest editors and external referees, the result is a section composed of 11 Hot Topic and six Big Trend articles.

Of course, this special section can only cover a small part of the breadth and depth of CS research and innovation pursued in Europe and Israel. Nevertheless, we hope it will provide insights into many of the strategic drivers and networks of what is happening in the region and into exciting research results and start-up initiatives.

We would like to thank Jakob Rehof, co-chair of *Communications'* Regional Special Sections Editorial Board, and Morgan Denlow, *Communications'* Deputy to the EIC, for their most helpful advice and support in preparing this section. Many thanks go to all submitters and presenters at the workshop for their valuable and constructive input, and especially to the authors by adhering to our extremely tight deadlines despite all the obstacles placed in our way by the COVID-19 pandemic. This section would not have been possible without any of you. 

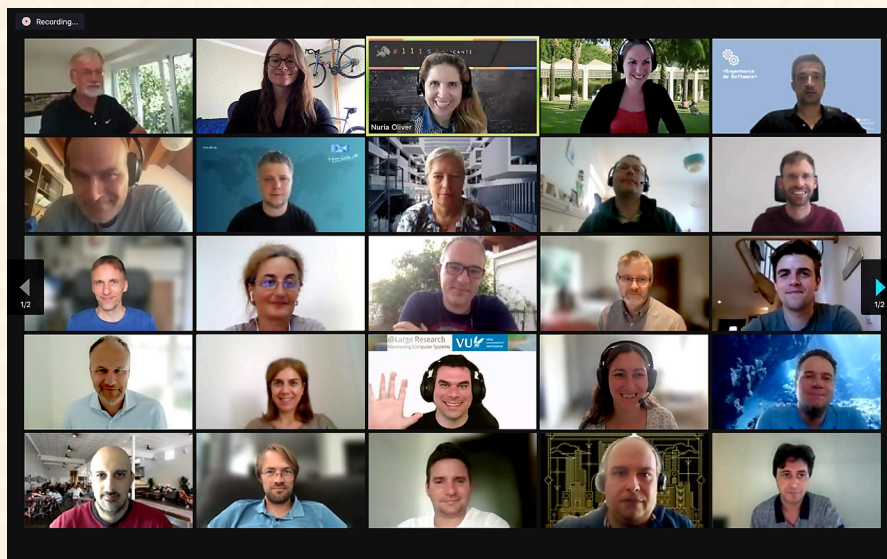
— **Jessica R. Cauchard, Matthias Jarke, and Nuria Oliver**
Europe Regional Special Section Co-Organizers

Jessica R. Cauchard is an assistant professor in Human-Computer and Human-Robot interaction in the Department of Industrial Engineering and Management at Ben Gurion University of the Negev, Israel.

Matthias Jarke is Emeritus Professor of Databases and Information Systems at RWTH Aachen University, Germany, and past Chairman of the Fraunhofer ICT group, the largest applied IT research organization in Europe.

Nuria Oliver is co-founder and Director of the ELLIS Unit Alicante Foundation and Chief Data Scientist at Data-Pop Alliance. She is also vice-president of ELLIS, the European Laboratory for Learning and Intelligent Systems.

Copyright held by authors.



More than 40 Europe Regional Special Section workshop participants met via Zoom to share ideas about the editorial content presented here.

EDITORIAL BOARD

EDITOR-IN-CHIEF

Andrew A. Chien
eic@cacm.acm.org

DEPUTY TO THE EDITOR-IN-CHIEF

Morgan Denlow
cacm.deputy.to.eic@gmail.com

CO-CHAIRS, REGIONAL SPECIAL SECTIONS

Jakob Rehof
Haibo Chen
P J Narayanan

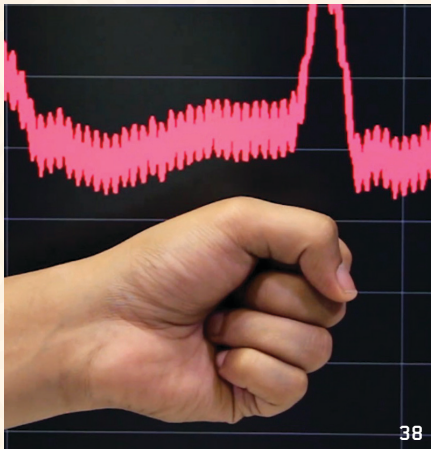
SPECIAL SECTION CO-ORGANIZERS

Jessica R. Cauchard
Ben Gurion University of the Negev, Israel
Matthias Jarke
RWTH Aachen University, Germany
Nuria Oliver
ELLIS Unit Alicante Foundation, Spain



Watch the co-organizers discuss this section in the exclusive *Communications* video.
<https://cacm.acm.org/videos/european-region-2022>

Hot Topics



38

36 **Trends in Computer Science Research within European Countries**
By Dima Kagan, Michael Fire, and Galit Fuhrmann Alpert

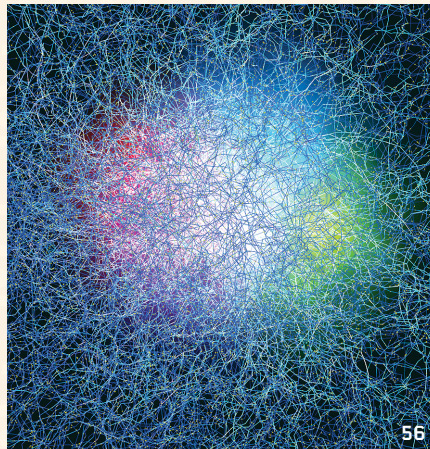
38 **On-Skin Computing**
By Jürgen Steimle

40 **Mobile Phone Usage Data for Disaster Response**
By Thomas R.C. Smallwood, Véronique Lefebvre, and Linus Bengtsson

42 **Robotic Process Automation Platform UiPath**
By Liliana Dobrica

44 **A Federated Infrastructure for European Data Spaces**
By Boris Otto

46 **Fenix: A Pan-European Federation of Supercomputing and Cloud e-Infrastructure Services**
By Sadaf R. Alam, Javier Bartolome, Michele Carpena, Kalle Happonen, Jacques-Charles Lafoucriere, and Dirk Pleiter



56

48 **On 6G and Trustworthiness**
By Gerhard P. Fettweis and Holger Boche

50 **Internet of Production—Entering Phase Two of Industry 4.0**
By Gertrude Kappel, Christian Brecher, Matthias Brockmann, and István Koren

52 **Privacy-Preserving AI for Future Networks**
By Diego Perino, Kleomenis Katevas, Andra Lutu, Eduard Marin, and Nicolas Kourtellis

54 **Partnership on AI, Data, and Robotics**
By Edward Curry, Fredrik Heintz, Morten Irgens, Arnold W.M. Smeulders, and Stefano Stramigioli

56 **Toward a Broad AIs**
By Sepp Hochreiter

Big Trends



74

58 **Energy Informatics—Key Elements for Tomorrow's Energy System**
By Hartmut Schmeck, Antonello Monti, and Veit Hagenmeyer

64 **Trust, Regulation, and Human-in-the-Loop AI**
By Stuart E. Middleton, Emmanuel Letouzé, Ali Hossaini, and Adriane Chapman

69 **Recommender Systems under European AI Regulations**
By Tommaso Di Noia, Nava Tintarev, Panagiota Fatourou, and Markus Schedl

74 **MobiGuide: Guiding Clinicians and Chronic Patients Anytime, Anywhere**
By Mor Peleg, Yuval Shahar, and Silvana Quaglini

80 **European Leadership in Process Management**
By Wil van der Aalst

84 **When Software Engineering Meets Quantum Computing**
By Shaukat Ali, Tao Yue, and Rui Abreu

Trends in Computer Science Research within European Countries

BY DIMA KAGAN, MICHAEL FIRE, AND GALIT FUHRMANN ALPERT

EUROPEAN INSTITUTES PROVIDE major worldwide contributions to research in multiple computer science (CS) domains. In 2020 alone, one million CS papers were published worldwide, over a third by European researchers. We used Microsoft Academic Graph,^a with nearly 5.5 million CS-related papers published by European institutes, across 34 CS subfields, to empirically evaluate European research and collaborations (see Figure 1). These findings

^a MAG is a bibliometric dataset containing over 263M publications.

provide value insight that policymakers could use in considering where to lead European CS when distributing budgets, by either encouraging leading fields and collaborations or strengthening those that fall behind.

In the last decade, 30% of worldwide CS publications were of European origin. For comparison, North America leads with 33% and Asia provides 30% of worldwide CS publications. Interestingly, in terms of worldwide attention, Europe also holds 30% of worldwide CS citations, while North America impressively approaches nearly half (47%) of them, Asia following third

with 16%. Other continents offer less than 6% contribution to total CS output and citations.

Inspecting dynamics reveals that some classically leading domains (for example, algorithms and telecommunications) are declining in favor of uprising domains (for example, data science, and human-computer interaction). Some domains (like computer vision) maintain stability over the years.

Breaking down contributions of individual countries within Europe, we explored both *volume*^b and *impact*^c of publications. Volume mapping reveals, for example, that papers in algorithms originate mostly from west European countries, most prominently Great Britain (GB) and Germany. Similar maps are also observed for all CS subfields collapsed.

^b Number of publications.
^c Number of citations.

Publication volumes are important, but do not necessarily imply quality or impact. One problem of increasing global concern is that publication volume has been dramatically increasing, affected by the *race to publish*. We therefore analyzed separately high/low impact publications (papers with fewer than five citations considered low impact).

The spatial distribution of citations is similar to volume mapping, with some smaller countries (like Switzerland and Netherlands) of high impact, and East-West impact gaps, possibly resulting from historical separation.

Since results may be affected by countries' size, reflected also by numbers of publishing institutes, we computed average citations per institute in each country.^d This approach

^d Number of citations, normalized by the number of publishing institutes per country.

In the last decade, 30% of worldwide CS publications were of European origin.

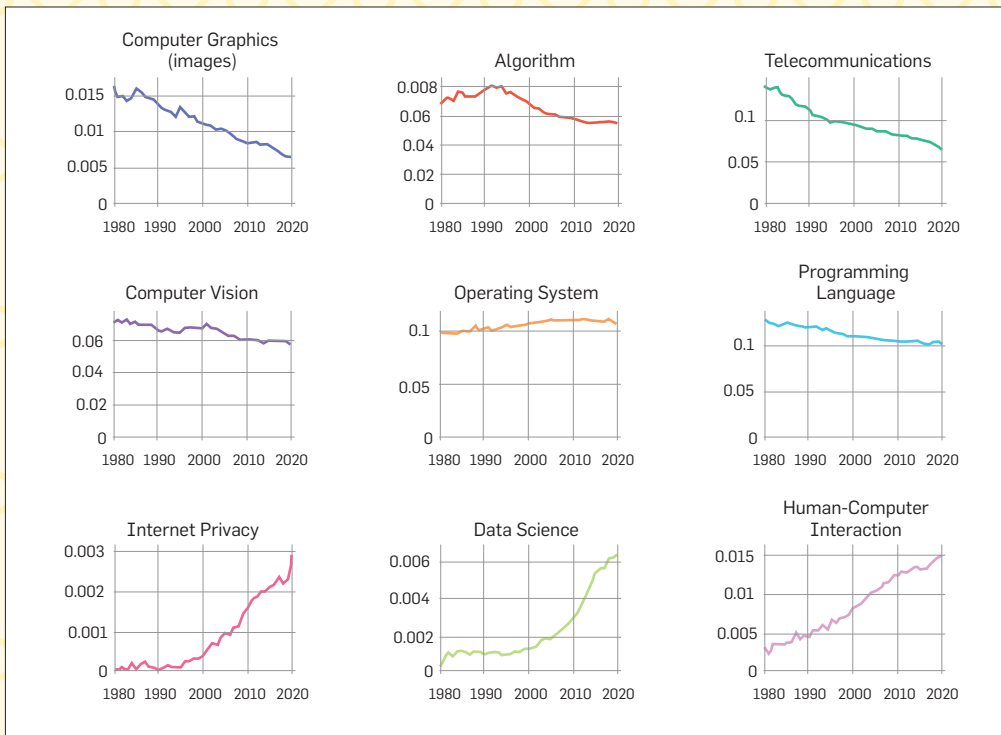


Figure 1. The number of European publications in different subfields, normalized by the total European CS publications.

highlights countries, including Switzerland, Italy, Netherlands, and Israel as leading citation rates per institute, despite not necessarily holding high country-wise volumes nor impact. Moreover, diving into subdomains highlights further specifics. For example, Israel contributes the highest institution citation rates in algorithms, followed by Italy.

Along with independent research, collaborations

are also essential for research. Understanding collaboration patterns may shed light on their importance for both promising funding and fruitful collaborations. We thus mapped volumes of collaborative publications.^e Russia and Poland exhibit the highest independent research, while other

^e For each country, percentage of published papers in collaboration with other European countries from all publications.

countries, like Belarus and Iceland, practice the highest ratios of collaborative research.

Mapping collaborative impact^f yields similar patterns, with some more collaborative countries (for example, Iceland and Belarus) benefiting from greater citation impact thanks to collaborations. Countries with low rates of collaboration can still benefit from collaborations, with more citations for collaborative research than publishing alone. East Europe, along with other less wealthy countries, stand out in this respect, suggesting that for world recognition collaborating with more acknowledged countries may be advantageous.

Benefits from research collaborations are not necessarily symmetric. Most

^f Percent citations of European collaborative papers with respect to all publications.

countries tend to benefit from collaborating, whereas some countries are better off publishing independently. For instance, GB and Switzerland have on average much less to gain from collaboration than Poland and Russia. East European countries generally have more to gain from collaborations than the rest of Europe.

Exploring all pairwise collaborations between European countries, in terms of give-take effective impact of collaborations allows one to choose promising collaboration partners. GB, for example, is better off publishing on its own than collaborating with most European countries.

CS research also expands to interdisciplinary fields. We quantified Europe's impact on interdisciplinary research, particularly in fields of growing interest related to health, energy, transportation, and aging. Of those studies, the highest European impacts are in health and public health, followed by energy related projects (as depicted in Figure 2).

To summarize, Europe has a significant role in global CS research. Careful inspection of leading fields, countries, and collaborations may support decision makers in distributing funds and shaping future European contribution to worldwide research, in the rapidly changing fields of computer science. 

Dima Kagan is a Ph.D. student at Ben-Gurion University of the Negev, Israel.

Michael Fire is an assistant professor in the Software and Information Systems Engineering Department at Ben-Gurion University of the Negev, Israel.

Galit Fuhrmann Alpert is an adjunct professor in the Software and Information Systems Engineering Department at Ben-Gurion University of the Negev, Israel.

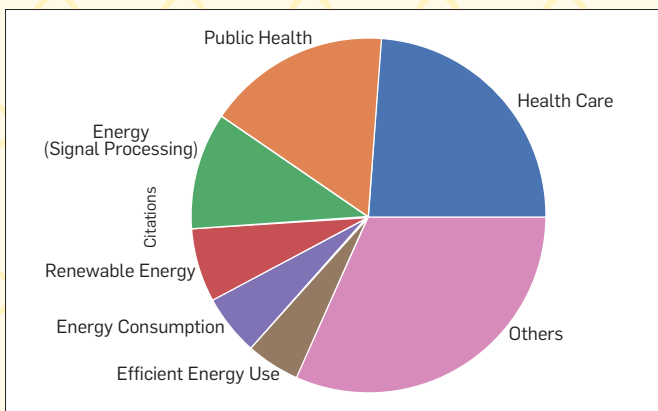


Figure 2. Percentage of CS papers in Europe related to the topics of health, energy, transportation, and aging.

On-Skin Computing

BY JÜRGEN STEIMLE

MODERN WEARABLE COMPUTERS are miniaturized, offer unprecedented mobility, and can even interface with the human body to monitor vital signs. Yet they have much more in common with their ancestor—the PC—than you might think. Just like old-fashioned computers, they are made of conventional electronics and therefore remain rigid and rather thick. This not only compromises ergonomics, but also limits the size of the devices and restricts where they can be deployed on the user. Can we instead make computers soft and malleable, such that they truly adapt to the human body?

A new generation of wearable devices, redesigned from the ground up, promises a significantly better compatibility with the human body. These so-called *epidermal devices*—devices modeled after human skin—are made of soft and stretchable materials and are two to three orders of magnitude thinner than traditional devices; in fact, they are typically much

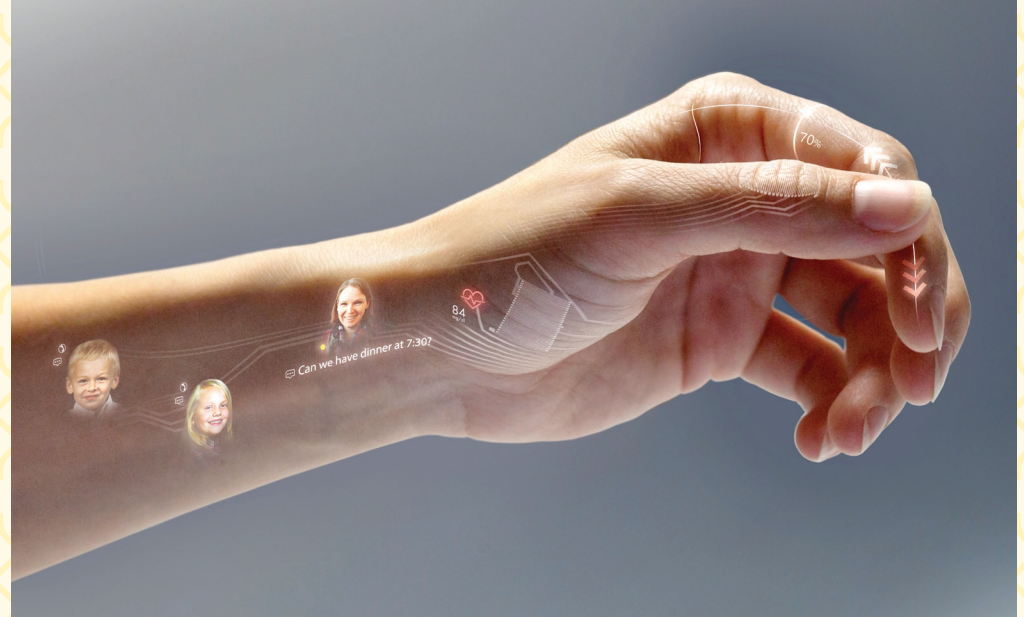


Figure 1. Future on-skin computing devices could seamlessly blend with the human body as visualized in this concept illustration of iSkin.⁷

thinner than the diameter of a single strand of hair.² Therefore, they can be worn as a barely noticeable patch on the skin. Embedded functional materials create fully flexible electronic components for sensing, output, processing, and power in a micron-thin form factor (see Figure 1).

The unique properties of skin have long fascinated researchers and have inspired electronic skin, soft sensors that mimic skin's mechanical properties while offering human-

like sensory capabilities.¹ Initially targeted at robots, electronic skins are increasingly being used on the human body. In recent years, a new multidisciplinary community has formed across the fields of new materials, electronics, biomedical engineering, and computer science that has made significant breakthroughs. This involves not only fundamental issues of materials and manufacturing, such as further improving minimally invasive wearability,⁴ but also functional designs for a wide range of applications.⁶

While early work focused primarily on monitoring biosignals, it soon became clear that devices on the skin offer principled new avenues for human-machine interfaces. We are investigating these at Saarländ University as part of

a five-year research project entitled “Interactive Skin,” funded by the European Research Council (ERC). Combining European strengths in human-computer interaction, graphics, AI, and specialized hardware technologies, our goal is to contribute to a new generation of devices that computationally augment the natural functions of our skin.

Merging Human Skin with Computational Augmentations

Controlling computing devices in demanding mobility conditions is a long-standing challenge, such as when the user's hands are busy holding an object or when the situation does not allow them to look at a screen. Devices worn on the skin offer a promising solution: a thin membrane with a touch

Can we make computers soft and malleable, such that they truly adapt to the human body?

sensor can be adapted to different parts of the body and provides an easy-to-reach surface for gestural input.⁷ For example, on a sensor placed on the index finger, users can make subtle touch gestures with their thumb, even while holding objects. Additional sensors to detect skin deformation and stretchable displays can also be integrated⁸ (see Figure 2 for examples).

Ultrathin devices on the skin can also lead to more natural experiences in augmented reality and virtual reality. To augment the use of everyday objects, we have presented a first-of-its-kind device for what we call feel-through haptics.⁹ The minimally invasive temporary tattoo is worn on the finger pad. It is so thin and soft that it even conforms to fine wrinkles, allowing users to feel real-world objects or surfaces they touch through the device. At the same time, the device augments real-world cues with high framerate electro-tactile output generated by eight densely spaced electrodes.


This enables novel forms of tactile augmented reality. For example, the system can change the way users perceive materials and objects, or it can create virtual haptic elements, such as virtual buttons, that the user can feel on an object even though they are not physically present.

Computational Design for Wearable Devices that Match the Human Body

Fitting devices to the anatomy of the human body unlocks exciting new opportunities, but it also makes the design of those devices more challenging. Not only does the device need to fit comfortably, but sensors may need to be positioned precisely on the body to detect body movement or pick up biosignals. This multifactorial design space makes it difficult even for experts to manually find optimal device designs. We see great potential for computational approaches to replace manual device design. In our recent work, we have shown that computational optimization with

We see great potential for computational approaches to replace manual device design.

anatomical models enables rapid design of highly compact physiological sensing devices that outperform expert-generated designs.⁵ The resulting designs can be printed on an office inkjet printer to create functional sensor tattoos.

First skin devices are already commercially available.³ However, many issues remain to be addressed before skin computers can enter mainstream. One important aspect relates to understanding social perceptions of using on-skin devices.¹⁰ In addition, the integration of processing, power supply, and networking into ultrathin and stretchable devices remains a formidable challenge that, once solved, could change computing and how it integrates with our body. 

References

- Hammock, M.L., Chortos, A., Tee, B.C.-K., Tok, J.B.-H., and Bao, Z. The 25th anniversary article—The evolution of electronic skin (e-skin): A brief history, design considerations, and recent progress. *Advanced Materials* 25, 42 (2013), 5997–6038.
- Kim, D.-H. et al. Epidermal electronics. *Science* 333, 6044 (2011), 838–843; <https://doi.org/10.1126/science.1206157>
- MC10 BioStamp; <https://www.mc10inc.com>. La Roche-Posay MyUV Patch; <https://www.lorealtechtincubator.com/myuupatch>.
- Miyamoto, A. et al. Inflammation-free, gas-permeable, lightweight, stretchable on-skin electronics with nanomeshes. *Nature Nanotech* 12 (2017), 907–913; <https://doi.org/10.1038/nnano.2017.125>
- Nittala, A.S., Khan, A., Karrenbauer, A., Kraus, T. and Steimle, J. Computational design and optimization of electrophysiological sensors. *Nature Communications* 12 (2021), 6351; <https://doi.org/10.1038/s41467-021-26442-1>
- Ray, T.R. et al. 2019. Bio-integrated wearable systems: A comprehensive review. *Chem. Rev.* 119, 8 (2019), 5461–5533; <https://doi.org/10.1021/acs.chemrev.8b00573>
- Weigel, M., Lu, T., Bailly, G., Oulasvirta, A., Majidi, C. and Steimle, J. iSkin: Flexible, stretchable and visually customizable on-body touch sensors for mobile computing. In *Proceedings of the 33rd Annual ACM Conf. Human Factors in Computing Systems*, 2015, 2991–3000. ACM, New York, NY; <https://doi.org/10.1145/2702123.2702391>
- Weigel, M., Nittala, A.S., Olwal, A., and Steimle, J. SkinMarks: Enabling interactions on body landmarks using conformal skin electronics. In *Proceedings of the 2017 CHI Conf. Human Factors in Computing Systems*, 3095–3105. ACM, New York, NY; <https://doi.org/10.1145/3025453.3025704>
- Withana, A., Groeger, D., and Steimle, J. Tacttoo: A thin and feel-through tattoo for on-skin tactile output. In *Proceedings of the 31st Annual ACM Symp. User Interface Software and Technology*, 2018, 365–378. ACM, New York, NY; <https://doi.org/10.1145/3242587.3242645>
- You, C.-W., Lin, Y.-F., Luo, E., Lin, H.-Y. and Kao, H.-L. Understanding social perceptions towards interacting with on-skin interfaces in public. In *Proceedings of the 23rd Intern. Symp. Wearable Computers*, 2019, 244–253. ACM, New York, NY; <https://doi.org/10.1145/3341163.3347751>

Jürgen Steimle is a professor of computer science at Saarland University, Saarland Informatics Campus, Saarbrücken, Germany.

 This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

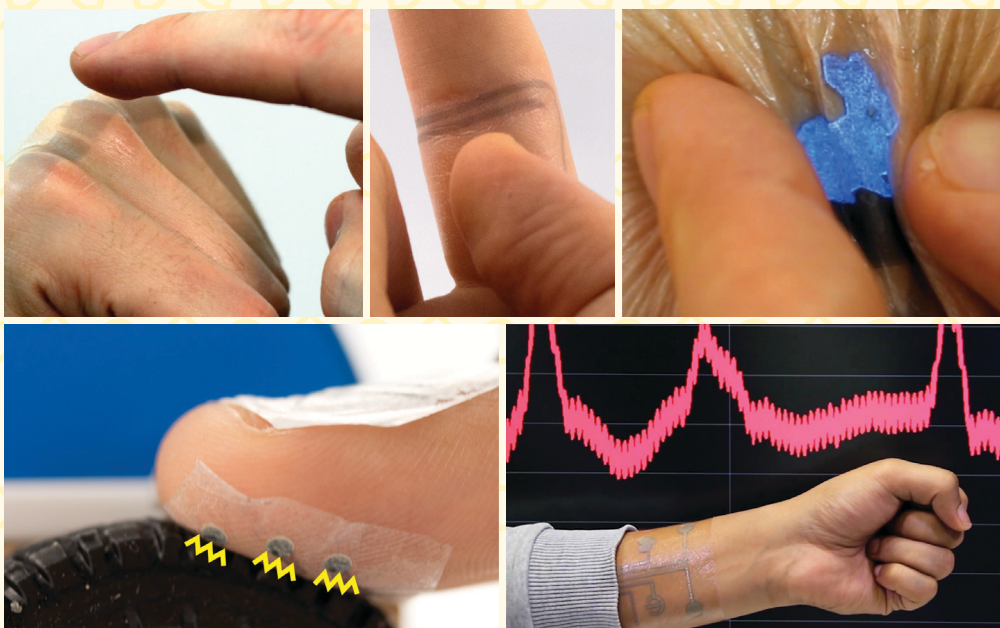


Figure 2. On-skin devices provide various functions, such as user input with rapid touch gestures, visual output, feel-through haptic feedback, and continuous monitoring of body vitals.

Mobile Phone Usage Data for Disaster Response

BY THOMAS R.C. SMALLWOOD, VÉRONIQUE LEFEBVRE, AND LINUS BENGTTSSON

SEVERE DISASTERS CAN CAUSE large population movements as affected people are displaced from their residences. The humanitarian response to such events relies on understanding where affected people are located. Mobility data can provide important insights at all stages of a crisis, from preparedness to long-term recovery.

At Flowminder, a non-profit foundation based in Sweden, the U.K., and Switzerland, we are working to support decision-makers to transform the lives of vulnerable people by facilitating access to novel sources of mobility data, including Call Detail Records (CDRs). Across Europe, non-profits, national statistical services, academic institutions, and mobile network operators (MNOs), including Orange, Telefonica, Telenor, and Vodafone, are collaborating on using CDR to support public policy. Additionally, the European Commission established the High-Level Expert Group on Business-to-Government Data Sharing to support such collaborations, including in response to the COVID-19 pandemic.⁹

The global spread of mobile devices provides new opportunities for better understanding mobility and addressing mobility data gaps. The International Telecommunication

Union estimates that 92% of the global population, including 89% of people in Least Developed Countries (LDCs), have mobile network coverage with 105 mobile-cellular telephone subscriptions per 100 people globally, and 74 subscriptions per 100 people in LDCs.⁵ However, mobile device penetration remains substantially lower in many vulnerable populations.

CDRs are routinely produced by MNOs for billing purposes each time a subscriber makes or receives a call, sends or receives a text message, or uses mobile data (“network events”). Each record contains the time of the event and

the cell site it was routed through, thereby describing subscribers’ movements. By pseudonymizing and aggregating CDR data over large numbers of subscribers, practitioners produce mobility indicators that provide insights into the overall movement of the population while preserving individual privacy. Furthermore, because this data is collected in near-real time, indicators can be calculated within days to aid the response to ongoing crises.

CDRs have been used to support responses to natural disasters in Low-to-Middle Income Countries (LMICs), including Bangladesh,⁸ Haiti^{1,7} (see Figure 1), Nepal¹²

and Vanuatu.⁶ It has a range of applications in disaster management, including dynamic population mapping and calculating flows between locations. These indicators reveal important information throughout a disaster, such as the number of people who may be affected by a disaster, the number complying with evacuation orders, the origins, and destinations of internally displaced people (IDPs), and the return of IDPs to their pre-disaster residence (see Figure 2).

Flowminder’s analysis of three disasters in Haiti and Nepal,² and confirmed for other disasters,¹³ has shown a consistent decay rate in the

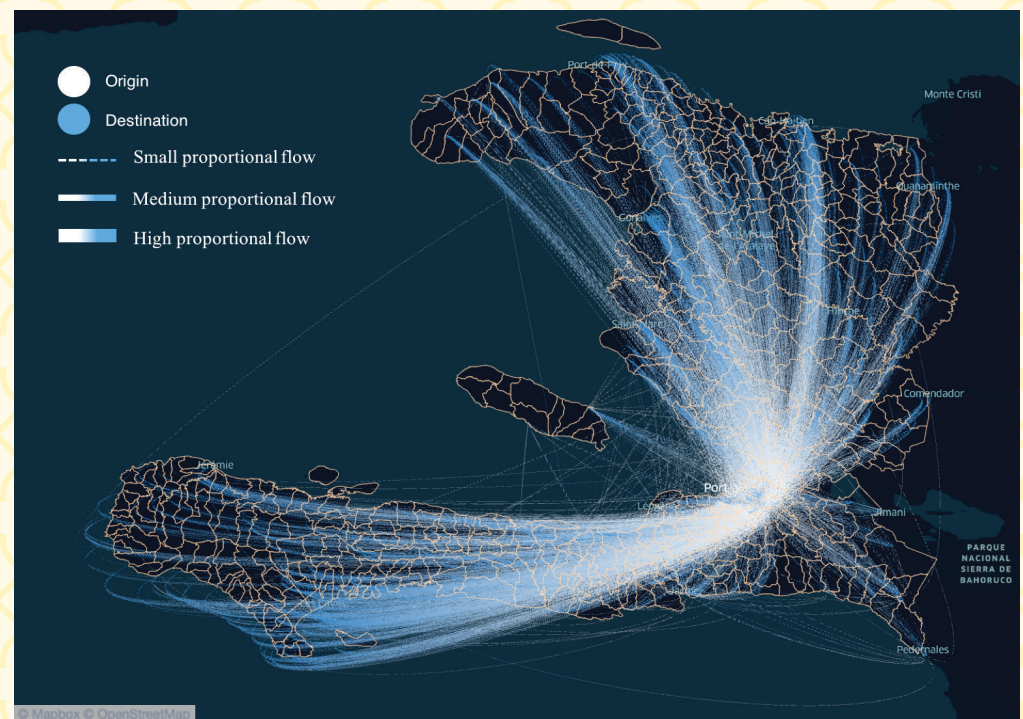


Figure 1. The displacement of people from Port-au-Prince in the aftermath of the 2010 Haiti earthquake (CDRs from Digicel Haiti).

number of IDPs remaining displaced post-disaster. This allowed Flowminder to forecast the remaining numbers of IDPs after the 2021 Haiti earthquake to help inform long-term planning.

CDRs can also improve predictions of displacement location. Modeling by Flowminder has shown that more localized travel and social contacts are associated with individuals being displaced in the vicinity of their pre-disaster residence, regardless of disaster intensity and local property damage.² Such models can support disaster preparedness, including simulations and exercises, and impact prediction.

Mobile phone subscribers are not, however, representative of the population, especially in LMICs. Subscriptions and usage vary with factors including gender, age, education, and socioeconomic status.¹¹ This is of particular concern for humanitarian use cases, compared to commercial applications, as it


may result in the impact of disasters on potentially vulnerable groups under-represented in the data being underestimated. Flowminder is developing methods for addressing these limitations, including gender-disaggregated indicators in Nepal³ and the collection of survey data in the Democratic Republic of Congo and Ghana.

Preserving individual privacy presents another challenge. CDRs contain the movement patterns of individual subscribers, though these are never shared. To ensure anonymity, Flowminder and other practitioners provide aggregates describing the mobility of large numbers of subscribers. While trajectories might be inferred from high-resolution aggregates, real-world datasets, and particularly those from LMICs, are resistant to re-identification attacks.¹⁰

Expediting the production of indicators to provide rapid, up-to-date informa-

CDR-derived mobility indicators can provide rapid, near real-time insights before, during, and after a disaster.

tion to decision-makers is also a key challenge. Partnerships in countries frequently impacted by natural disasters, such as between the GSMA and TuckCell in Turkey,⁴ can facilitate this as access to data has already been agreed and implemented. Following the 2021 earthquake, Flowminder's existing partnership with Digicel Haiti enabled the release of a report just six days after the earthquake. Flowminder is also supporting MNOs to establish streamlined technical and administrative processes through projects like FlowKit and OPAL, to facilitate the timely production of mobility indicators.

CDR-derived mobility indicators can provide rapid, near real-time insights before, during, and after a disaster, when understanding population movements and mobility is important to the efficient provision of aid. There remain important areas for further improvement of these tools to expedite the dissemination of the indicators and to address the biases in CDR data. Regardless, such indicators are immensely valuable to decision making, especially during a disaster when high-quality data is limited. 

References


1. Flowminder. Hurricane Matthew Haiti estimated population movement as of 22 November 2016; <https://bit.ly/3Kua7sd>.
2. Flowminder. Contributing to a better understanding of human mobility in crisis: Technical Report. 2019. Available on request.

3. Flowminder. Towards high-resolution sex-disaggregated dynamic mapping, 2019; <https://bit.ly/3KqIjVW>
4. GSMA. Big data for social good: Utilising real-time mobile analytics to inform emergency disaster response in Turkey, 2019; <https://bit.ly/3Ag04CF>
5. International Telecommunication Union. Measuring digital development: Facts and figures 2020. ITU, Geneva, Switzerland.
6. Khaefi, M.R., Prahara, P.J., Rhea, M., Alkarisyia D. and Hodge G. Predicting evacuation destinations due to a natural hazard using mobile network data. In *Proceedings of the 2nd Intern. Conf. Informatics and Computational Sciences*, 2018, 1–6.
7. Lu, X., Bengtsson, L. and Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. In *Proceedings of the National Academy of Sciences* 109, 29 (2012), 11576–11581.
8. Lu, X. et al. Detecting climate adaptation with mobile network data in Bangladesh: anomalies in communication, mobility and consumption patterns during cyclone Mahasen. *Climatic Change* 138 (2016), 505–519.
9. Vespe, M., Iacus, S., Santamaria, C., Sermi, F. and Spyros, S. On the use of data from multiple mobile network operators in Europe to fight COVID-19. *Data & Policy* 3 (2021).
10. Wang, H., Gao, C., Li, Y., Wang, G., Jin, D. and Sun, J. De-anonymization of mobility trajectories: Dissecting the gaps between theory and practice. In *Proceedings of the 2018 Network and Distributed Systems Security Symp.*
11. Wesolowski, A., Eagle, N., Noor, A.M., Snow, R.W. and Buckee, C.O. The impact of biases in mobile phone ownership on estimates of human mobility. *J. Royal Society Interface* 10, 81 (2013).
12. Wilson, R. et al. Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake. *PLoS Currents* 8 (2016).
13. Yabe, T., Tsubouchi, K., Fujiwara, N., Sekimoto, Y. and Ukkusuri S.V. Understanding post-disaster population recovery patterns. *J. Royal Society Interface* 17, 163 (2020).

Thomas R.C. Smallwood is Knowledge Centre Manager at the Flowminder Foundation, Southampton, U.K.

Véronique Lefebvre is Director of Data Analysis at the Flowminder Foundation, Southampton, U.K.

Linus Bengtsson is Chair of the Board for the Flowminder Foundation, Stockholm, Sweden.

 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. <http://creativecommons.org/licenses/by/4.0/>

Preparedness (Baseline)	How many people live within phone coverage areas?
	How many active subscribers are there per region per month?
	How many people could be reached via mobile phones?
Hazard forecast	How many people may be affected?
Hazard warning	Are people moving or relocating as the results of warnings?
Active crisis with network outage	Have cell towers been damaged?
Active crisis	How many subscribers are displaced, from where to where?
Short-term recovery	Are subscribers returning to their home or staying at their displaced locations, and how long does it take (in terms of weeks)?
Long-term recovery	Are resettlement areas growing or depleting over time (in terms of months, years)?

Figure 2. The applications of CDR-derived mobility indicators over the course of a humanitarian crisis.

Robotic Process Automation Platform UiPath

BY LILIANA DOBRICA

THE UIPATH PLATFORM combines core robotic process automation (RPA) capabilities with tools for process discovery and analytics to report precisely the business impact. The core capabilities make it easy to build, deploy, and manage software robots (SRs) that emulate humans' interactions with information systems to perform certain tasks in business processes (BPs). Firstly, the BPs to be automated are designed, created, or recorded. They are created using drag-and-drop activities within a workflow. Then SRs work to perform BPs and an orchestrator acting as a control center designates tasks/processes to SRs and evaluates the efficiency of each one. A tool takes screenshots with every mouse click or keyboard input and collects smart data about process statistics such as execution time or number of actions. SRs follow the choreog-

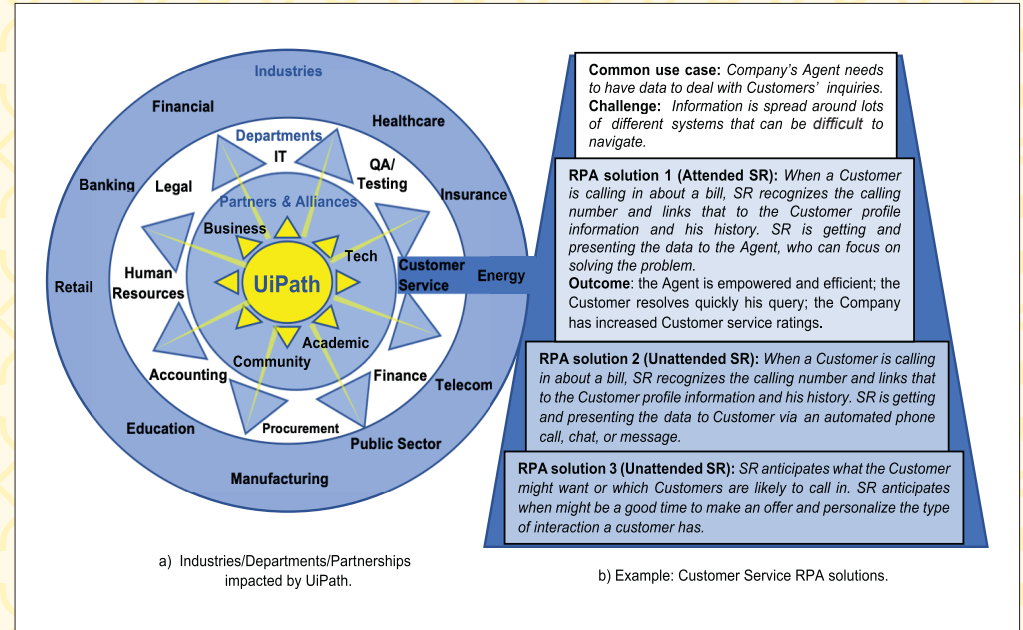


Figure 1. UiPath impact.

raphy of specific modules, while operating with IT infrastructure and using established applications. SRs can understand what is on a screen, complete the right keystrokes, navigate systems, identify and extract data, and perform a wide range of defined actions. Advanced SRs can perform cognitive

processes like interpreting text, engaging in chats and conversations, understanding unstructured data, and making complex decisions by applying advanced ML models.

The current challenges with the UiPath platform include the need for continuous releases to increase value through the integration of more cognitive and cybersecurity services. It encourages new guidelines and standards that guarantee trust and safety use. Academics help educate potential adopters by objectively researching actual implementations, by assessing them and extracting lessons on realizing value. A common effort with academics is essential to develop harmonization.

Steps toward consensus include the definition of a clear and consistent terminology in the use of concepts that belong to the RPA, a taxonomy of product features and functionality, or some recommended practices for implementation and management methodologies.

European interest is increasing digital innovation by promoting start-ups, scale-ups, and unicorns and supporting their inclusion in the ecosystems with a direct impact on the growth of global success. Being supernatural creatures, unicorns are very rare. Their friendliest places to develop provide a financial and innovation-driven climate. Eastern Europe has some of the

UiPath is a remarkable European success exported to the world in the RPA space to digitalize societies and to build a complex ecosystem driven by innovation.

best and competitive ICT talents and the desire to connect to the world. Particularly, Romania has the education strength in computer science and mathematics, but the culture of entrepreneurship is young and it takes time to grow. The founders and the platform R&D team of UiPath are computer science graduates from Romanian universities. In the new reality related to the digital transformation,¹ Romania has developed into a center of excellence becoming an outsourcing regional hub in ICT.


UiPath is a remarkable European success exported to the world in the RPA space to digitalize societies and to build a complex ecosystem driven by innovation (see Figure 1a). It contributes with RPA solutions⁶ to current challenges by creating virtual digital workforce and solving social issues. Many industries are worried about the decrease of the workforce in the near future and face major problems due to ageing or overworked employees in departments that depend on best usage of digital data. SRs can automatically execute deterministic, repetitive, standardized, high volume, and rules-based tasks by capturing and analyzing structured data and working across several interoperable systems. SRs are best suited to do swivel chair-data entry, taking information from one system and integrating it with another. Figure 1b provides several RPA solutions to a common use case challenge.

UiPath is an important software platform provider with Romanian roots that became a successful global tech company with head-

quarters in the U.S., where the deepest ecosystems are located. It grew explosively, both in size and value, to a unicorn and recently scaled up to a decacorn. An article in *The Economist*² declared UiPath as “Europe’s most successful tech export since Spotify” and estimates it holds a third of the market. Moreover, ranking analyzes from Gartner^{3,5} places it the highest position and recognizes UiPath as the “2021 Magic Quadrant Leader in the RPA space” for its ability to execute for the third consecutive year.

The fantastic journey of the Romanian start-up that began in 2005 was marked on April 21, 2021—a glorious day for UiPath as it

became a decacorn and a publicly trading company on the New York Stock Exchange. Due to an insufficient entrepreneurial climate, the path toward success wasn’t straightforward⁵ (Figure 2). Only the team value, the software platform quality, and the commitment to building a large company convinced investors, clients, and business partners. Its great innovation occurred during the rise of AI and cloud technologies when its growth took off with a new product flightmap that integrates ML and AI algorithms within customers’ business operations. In 2018, it was the fastest-growing enterprise soft-

ware company ever with the most widely used RPA platform in the world. Since then it has remained in the leader position. 

References

1. Bucharest 5A+ Choice for Cyber. Oct. 2020; <https://bit.ly/3HkUO2Q>
2. Business Software: The robots are coming. *The Economist*, (Apr 24, 2021); <https://econ.st/3pCKpcQ>
3. Gartner peerInsights. UiPath Platform Ratings Overview; <https://gtrn.it/319XuRk>
4. Gheorghe, G. The story of UiPath—How did it become Romania’s first unicorn. *Business Review* (2018); <https://bit.ly/3pDyHhY>
5. UiPath. Gartner Magic Quadrant for Robotic Process Automation, 2021; <https://bit.ly/3pEZB9n>
6. UiPath. Automation Case Studies, 2021; <https://www.uipath.com/resources/automation-case-studies>

Liliana Dobrica is a professor in the Faculty of Automation and Computer Science at the University Politehnica of Bucharest, Romania.

© 2022 ACM 0001-0782/22/4 \$15.00



Figure 2. Recognition race—pioneering work, timeline of some key UiPath events.

A Federated Infrastructure for European Data Spaces

BY BORIS OTTO

THE EUROPEAN STRATEGY FOR DATA CALLS for “common data spaces” as a foundation for the data economy in the European Union.¹ President Ursula von der Leyen expressed her view that Gaia-X should play a key role in response to these calls by providing a federated data infrastructure in the cloud.^a

One data space example is the Mobility Data Space^b that was launched in 2020 as a multistakeholder project in Germany in close alignment with the German Federal Government. It aims at data-driven services in the mobility sector and data sovereignty of the data holders and trust among all participants. Smart service examples are intermodal

end-to-end services for individual travelers, traffic management services for smart cities, and services to increase road safety for individual drivers.

Figure 1 shows the basic functionality of the Mobility Data Space following the provisions of the International Data Spaces (IDS) Reference Architecture Model (RAM).⁴ It does not pool data in a central data store but rather connects data providers and data users by the respective use of the IDS Connector component. A catalog allows data providers to present and describe their data resources, together with conditions under which the data can be used. Data users search the catalog for data they need for their smart service. If data demand and supply match, the exchange of the data itself is carried out just between the participants, with no involvement of the

Mobility Data Space operator. Only metadata on data exchange transactions are monitored and logged to ensure conditions for the use of the data are correctly exchanged and followed.

The Mobility Data Space balances the interest of the individual data provider regarding data sovereignty and the interest of the community to increase the data availability and its use. Data has a value when used⁷ and, thus, should be used as widely as possible to seize its innovation opportunities. However, data use must always take into consideration the interests of the data holder. Providing data, particularly high-quality data, comes at a cost.⁶ In this context, data sovereignty is defined as the capability of a data holder to be self-determined regarding the use of their data.⁵

Gaia-X as a Federated Data Infrastructure

Data spaces represent a data integration concept that follows Linked Data design principles.² First, data spaces do not require physical data integration but leave the data at the source and only make it accessible when needed. Second, data is not forced into one common schema but linked, that is, integrated on a semantic level. Third, the distributed architecture of data spaces allows data redundancies, that is, multiple data objects may exist

in a data space describing the same real-world object. Finally, data spaces can be overlapping and nested so that data holders and data users, respectively, can be participants in multiple data spaces.

To implement the European Strategy for Data and create common data spaces, interoperability of data and collaboration of participants is required across the boundaries of individual data spaces. Gaia-X aims to achieve this by providing so-called federation services that function within and across different data spaces.

The Gaia-X initiative is organized as a not-for-profit association headquartered in Brussels.^c With more than 300 members, the association aims at a federated data architecture that comprises not only data and smart services, but also cloud infrastructure services (see Figure 2). Gaia-X specifies four so-called federation services, namely “identity and trust,” “sovereign data exchange,” “federated catalog,” and “compliance.”³ They form a blueprint for data spaces and allow for “federations of ecosystems,” hence, support interoperability and coordination across data spaces. The latter requires federation of identities of data space participants, of catalog entries, and of data transaction logs.

a See https://ec.europa.eu/commission/presscorner/detail/ov/SPEECH_20_1655

b See <https://mobility-dataspace.eu/>

c See <https://www.gaia-x.eu/>

Gaia-X supports the goals of interoperability of services, data portability, data sovereignty as articulated in the European data strategy, and addresses a requirement gap in existing infrastructure models.

Data Infrastructure Patterns

Gaia-X supports the goals of interoperability of services, data portability, data sovereignty as articulated in the European data strategy and addresses a requirement gap in existing infrastructure models. The latter were—and still are—dominated by hyper-scaling platform providers on the one hand side and state-controlled infrastructures on the other hand. Both do not meet the European requirements mentioned here. Consequently, Gaia-X represents an alternative design pattern for data infrastructures.

In contrast to alternative patterns, Gaia-X follows a cooperative approach (see approximate juxtaposition in the accompanying table). The federated infrastructure is open to be used, owned by the community itself, and thus avoids concentration of control over data and services. **C**

References

1. European Commission. A European strategy for data. Brussels, Belgium, 2020.
2. Franklin, M., Halevy, A., Maier, D. From databases to dataspace. *ACM SIGMOD Record* 34, 4 (2005), 27–33; <https://doi.org/10.1145/1107499.1107502>.
3. Gaia-X. Gaia-X Architecture Document (21.09 Release). Gaia-X European Association for Data and Cloud AISBL, 2021, Brussels, Belgium; <https://bit.ly/3sJkUd>
4. IDS Association. Reference Architecture Model (Ver. 3.0). International Data Spaces Association, Berlin, Germany, 2019; <https://internationaldataspaces.org/download/16630/>.
5. Jarke, M., Otto, B., Ram, S. Data sovereignty and data space ecosystems. *Business & Information Systems Engineering* 61, 5 (2019), 549–550; <https://doi.org/10.1007/s12599-019-00614-2>.
6. Redman, T.C. The impact of poor data quality on the typical enterprise. *Commun. ACM* 41, 2 (Feb. 1998), 79–82; <https://doi.org/10.1145/269012.269025>.
7. Tayi, G.K. Ballou, D.P. Examining data quality. *Commun. ACM* 41, 2 (Feb. 1998), 54–57; <https://doi.org/10.1145/269012.269021>.

Boris Otto is Chair for Industrial Information Management at TU Dortmund University and Director of the Fraunhofer Institute for Software and Systems Engineering in Dortmund, Germany.

Copyright held by author/owner. Publication rights licensed to ACM.

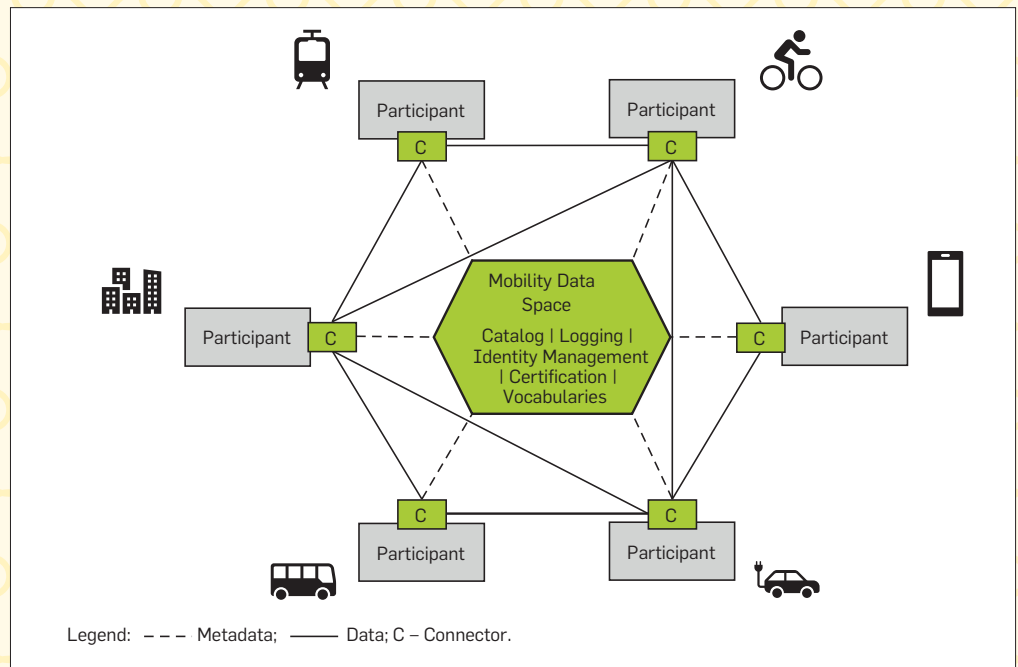


Figure 1. Mobility data space overview.

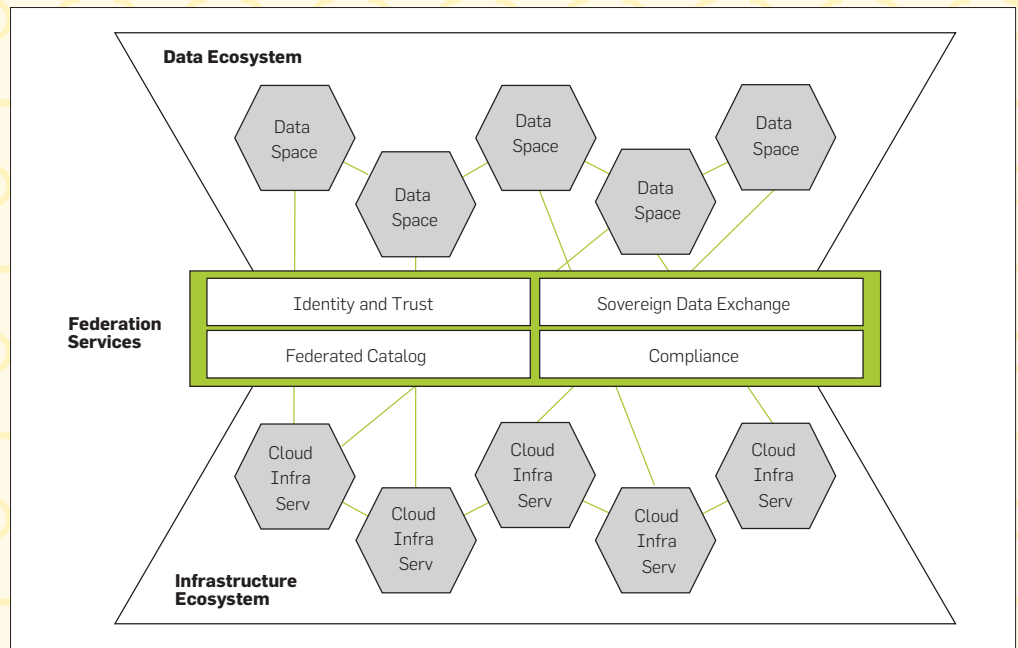


Figure 2. Gaia-X ecosystem and federation services.

	«Hyperscaler»	State-Controlled	Gaia-X
Economic Effect	“Winner takes all”	Digiriste	Cooperative
Platform Owner	1	1	Many
Platform Architecture Design	Central	Central	Federated
Platform Development	Closed/Hybrid	Closed	Open
Code Basis	Proprietary and Open	Proprietary and Open	Open
Data Sovereignty	Basic	Rudimentary	Core Value Proposition
Data Management and Data Exchange	Central	Central	Federated and Bilateral

Data infrastructure patterns.

Fenix: A Pan-European Federation of Supercomputing and Cloud e-Infrastructure Services

BY SADAF R. ALAM, JAVIER BARTOLOME, MICHELE CARPENE, KALLE HAPPONEN, JACQUES-CHARLES LAFOUCRIERE, AND DIRK PLEITER

TO ADDRESS INCREASING demands of various research communities for computing and storage services, six leading European supercomputing centers began harmonizing and federating their e-infrastructure services portfolio with the goal of supporting a variety of science and engineering communities. Barcelona Supercomputing Centre (BSC) in Spain, France's Commissariat à l'énergie atomique et aux énergies alternatives (CEA), Italy's Supercomputing Centre (CINECA), Finland's Supercomputing Centre (CSC), the Swiss National Supercomputing Centre (CSCS), and Juelich Supercomputing Centre (JSC) in Germany, have aligned high-end computing and storage services to facilitate the creation of the Fenix Research Infrastructure, which has been making resources available at scale to research communities since 2018.²

Characterized by different types of data repositories, scalable supercomputing systems, and private cloud instances, the Fenix portfolio is complemented by a federated identity and access management system.³ Presently, a diverse

portfolio of services are available for HPC, AI, and ML, and cloud computing applications, free of charge (<https://fenix-ri.eu/access>). Evaluation of the applications follows the peer review principles established by PRACE (<https://prace-ri.eu/>). The Fenix objective is to serve science and engineering domains that strongly benefit from diverse e-infrastructure services for their collaborative research and data sharing. It therefore leverages national, European, and international funding programs to realize the compute, storage, and network resources sustaining the e-infrastructure services. There are similar national programs such as U.S. NSF XSEDE (<https://www.xsede.org/>). However, Fenix introduces unique aspects: First, it defines a federated research e-infrastructure architecture for leadership-class supercomputing resources providers, transcending national boundaries; and second, it offers a uniform, federated identity, and access management solution.

A Co-Designed Solution for Science and Scientists

Development of Fenix services and the underlying technical solutions has been an iterative, co-design

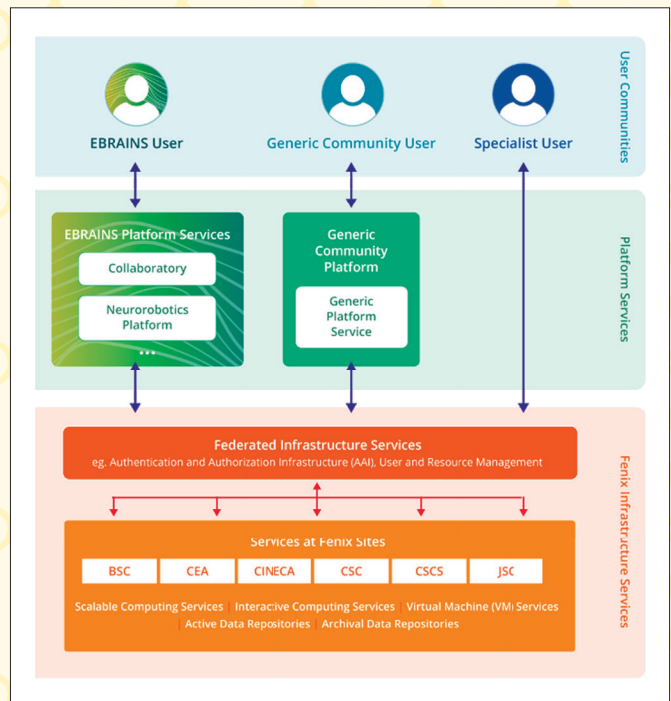


Figure 1. The Fenix consortium, and its federated and e-Infrastructure services portfolio.

process, which was initially driven by the Human Brain Project (HBP)—a flagship venture currently funded by the European Commission for a period of 10 years. Fenix has been facilitating the design, implementation, and operations of domain-specific (neuroscience) platform services. The need to federate services arose from a collaboration between scientists working on, for example, a Brain Atlas, an HBP platform service that requires integration of data from a variety of research teams

throughout Europe, and beyond.⁴ Various neuroscience workflows have in common the need of being able to collect data at the edge (that is, instruments running measurements), moving it to a nearby or affiliated datacenter, and making it available for further compute-intensive processing and integration with other datasets. The latter may come from geographically distributed datacenters.


Figure 1 highlights the concept of Fenix that has been co-designed with neu-

rosience and other similar use cases, for instance, the Materials Cloud, which is a platform designed to enable open and seamless sharing of resources for applications in materials modeling, exploiting supercomputing and cloud computing resources.⁵ The initial instantiation of Fenix is funded under a specific grant agreement of the HBP named the Interactive Computing E-Infrastructure (ICEI). Breakthrough scientific research exploiting Fenix resources has been documented at <https://fenix-ri.eu/infrastructures/success-stories>. For instance, an open source platform for constructing and simulating personalized brain network models using Fenix resources supports The Virtual Brain (TVB) workflows. The successful reconstruction and simulation of the cerebellar neurons and networks is another example. Both TVB and the Cerebellar Modelling Hub are part of a digital research infrastructure called EBRAINS, which has been included in the

2021 Roadmap of the European Strategy Forum on Research Infrastructures (ESFRI). The Fenix storage services have been used to share results of SARS-CoV-2 virus investigations. These projects highlight that Fenix resources not only enable access to HPC and cloud resources but are also on the critical path for realizing sustainable, digitalized research platforms for diverse scientific communities.

Federated Identity, Access, and Resource Management

Fenix leverages the European Authentication and Authorisation for Research and Collaboration (AARC) project's blueprint architecture for establishing federated identity and access management services.¹ As shown in Figure 2, the central proxy service is provided by GÉANT that manages one of the largest academic and research networks. The solution offers multiple levels of assurance and trust across the hosting sites as Identify Providers (IdPs) and communities such as

the HBP IdP. Fenix User and Resource Management Services (FURMS) provides federated access management to HPC and cloud resources. The core objectives of Fenix federation are a uniform experience for users, and extensibility such that Fenix AAI can be leveraged by the community or domain-specific platform development teams transparently. Fenix AAI facilitates identification and authentication of users by federating multiple IdPs, validating user profiles, maintaining a registry of usage agreement and policies, and the general Fenix usage agreement. FURMS provides central accounting, budgeting, and reporting mechanisms at different granularities (research groups or communities) and offers secure, role-based access controls. Furthermore, it serves as an attributes provider needed for authorization of Fenix services, for instance, secure ssh key management for HPC. 

References

1. AARC Community members, & AppInt members. AARC Blueprint Architecture 2019 (AARC-G045).

2. Zenodo; <https://doi.org/10.5281/zenodo.3672785>.
3. Alam S. et al. Fenix: Distributed e-infrastructure services for EBRAINS. Brain-Inspired Computing. K. Amunts, L. Grandinetti, T. Lippert and N. Petkov N., Eds. *Lecture Notes in Computer Science 12339* (2021). Springer, Cham; https://doi.org/10.1007/978-3-030-82427-3_6
4. Alam S.R. et al. Archival data repository services to enable HPC and cloud workflows in a federated research e-infrastructure. In *Proceedings of the 2020 IEEE/ACM Intern. Workshop on Interoperability of Supercomputing and Cloud Tech.*
5. Amunts, K., Mohlberg, H., Bludau, S., Zilles, Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science* (2020), New York, NY.
6. Talirz, L. et al. Materials Cloud, a platform for open computational science. *Sci Data* 7, 299 (2020); <https://doi.org/10.1038/s41597-020-00637-5>

Sadaf R. Alam is Chief Technology Officer of the Swiss National Supercomputing Centre (CSCS) at ETH Zürich, Switzerland.


Javier Bartolome is Systems Group Manager at the Barcelona Supercomputing Center (BSC), Spain.

Michele Carpenè is Software Developer and Project Manager at the Italian Supercomputing Centre (CINECA), Italy.

Kalle Happonen is Development Manager at the Finnish Supercomputing Center (CSC), Espoo, Finland.

Jacques-Charles Lafoucrière is Program Manager at the Commissariat à l'Énergie atomique et aux énergies alternatives (CEA), France.

Dirk Pleiter is a professor of computer science at KTH and Director of PDC, Sweden and Senior Researcher, Juelich Supercomputing Centre (JSC), Germany.

 This work is licensed under a Creative Commons Attribution 4.0 International License. <http://creativecommons.org/licenses/by/4.0/>

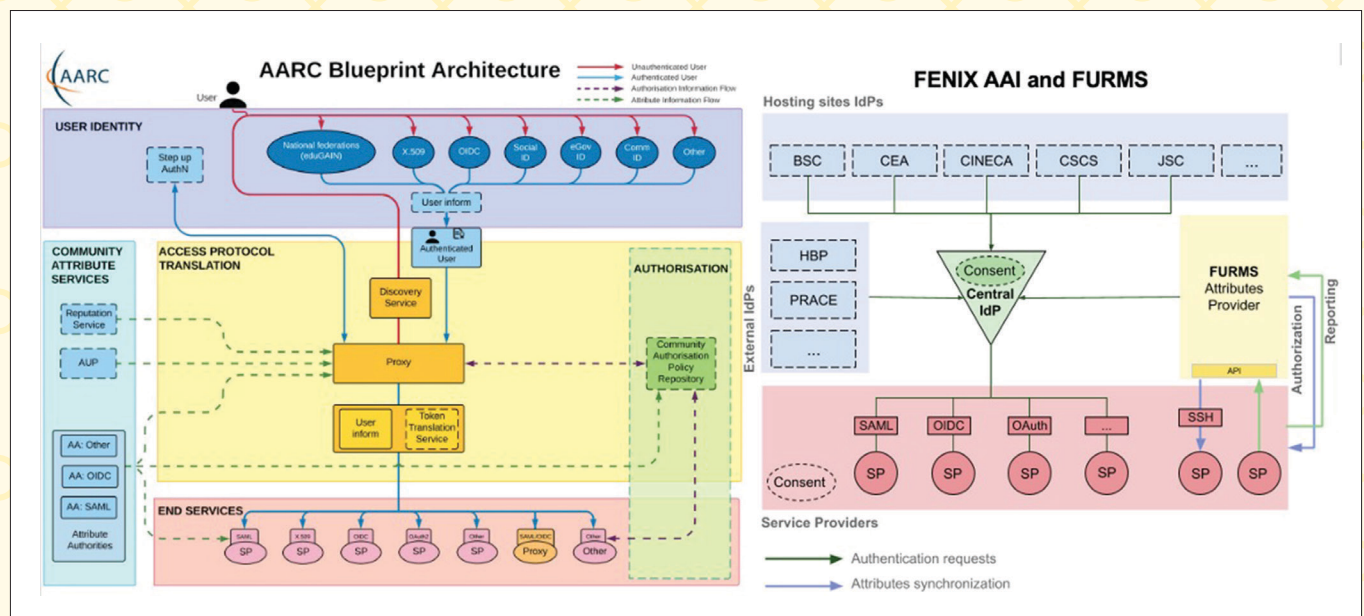


Figure 2. Fenix AAI and FURMS implementation compliant with AARC Blueprint Architecture.

On 6G and Trustworthiness

BY GERHARD P. FETTWEIS AND HOLGER BOCHE

THE FIRST TWO generations of cellular—1G/2G—enabled ubiquitous voice connectivity. 3G/4G enabled broadband Internet. Even generations introduced services for business customers, and odd generations democratized them for consumers. 5G is enabling network-controlled robotics and XR, the Tactile Internet for business verticals,⁴ and 6G will democratize this for consumers. One main avenue for achieving this is cost reduction.⁶ Another avenue is radio access with joint communications and sensing.⁷ New services are envisioned, such as low-altitude air traffic control, detecting, for example, bird migration and adapting drone services accordingly.

Not only data but physical and virtual objects will be controlled with 6G. This requires addressing *trustworthiness* of the system and its services at an unprecedented level. Indeed, trustworthiness must be understood in a new context, as we envision:

Every opportunity of improving sensing is an opportunity for spying. Trustworthiness for 6G is key.



The trustworthiness of 6G technology has crucial implication. The University of Oulu in Finland recently acquired a self-driving car from Toyota to be used as a piece of research equipment where researchers can install their own instruments for testing.

- ▶ Localization of unheard precision,
- ▶ Sensing—not only radio and camera sensing, and
- ▶ Gesture recognition—also emotions.

How can we provide these new qualities without compromising legal and societal requirements, for example, General Data Protection Regulation (GDPR)? Every opportunity of improving sensing is an opportunity for *spying*. Trustworthiness for 6G is key. It comprises:

- ▶ Privacy
- ▶ Security
- ▶ Integrity
- ▶ Resilience
- ▶ Reliability
- ▶ Availability
- ▶ Accountability
- ▶ Authenticity
- ▶ Device independence

Mathematical Frameworks

For communication tasks beyond Shannon's theory for message transmission, like event-driven-communication, transmission of status states, and joint communication and sensing, we must develop a Post-Shannon information theory. Several Post-Shannon transmission and storage schemes achieve exponential gains compared to the Shannon and Turing approaches.^{2,6} Besides, initial Post-Shannon transmission methods allow a secure transmission of information, which

cannot be broken even by quantum computers of arbitrary complexity. One important feature of 6G is resilience by design. This is particularly interesting since the successful execution of jamming attacks by an attacker cannot be detected by Turing machines.⁶

However, we must not only design systems that are robust against attacks from the outside, but also from within. Many cryptographic tasks have emerged in the last decade. Important examples are oblivious transfer, secure computing, bit commitment, and information masking. These tasks involve two or more untrusted parties with different types of behavior.¹ Some of the parties may be dishonest or even jam the communication system. It is well known that oblivious transfer is the most

powerful cryptographic two-party primitive.

We must develop new information theoretic tools to achieve oblivious transfer, secure computation, and information masking under real-world communication conditions.¹⁰ Combining quantum communication with classical communication offers additional advantages. It is an interesting research question if one can combine tools from quantum information theory like entanglement with classical tools from the theory of zero-knowledge proofs to achieve device and hardware independent trustworthiness.

Platforms for Trustworthiness

Building a computing platform for trustworthiness poses enormous challenges; new tools are required, as previously noted. Some major ones are:

1. The hardware/operating system platform must be trustworthy. Today's separate design must change to an integrated approach.³
2. Isolation ("barrier skin"⁵) must guarantee GDPR conformance, for any (cloud) services.
3. With increasing sensing

capabilities of terminals, the raw sensing data must be isolated (for example, encrypted¹¹) at the source. Specialized processing containers in the edge or in terminals will ensure, for example, that identity and location are only accessed by approved services. The orchestration could be carried out by a meta operating system.¹²

4. Integrity will be a major issue. Current cellular standards are written in English text—not machine readable. 6G must be specified in an ontology that allows formal verification and cross-checks of implementation code, including updates. This ensures trust in cellular infrastructure adhering to specification.⁹

5. Reliability and availability will not only be a challenge to be served at the network layer, but also at the radio layer.⁸

6. Network resilience has always been key for telecommunications. But this must be extended to the Radio Access Network as 6G will control mobile robots, including classical steps: monitor, respond, and counter.


7. An open question regards device independence: Can this be addressed in the context of trustworthi-

The step from 5G to 6G is not small, in fact, it's huge; as the vision of 6G enabling personal mobile robotics and XR requires far more than an "update."

ness, or must standardized platforms be adopted?

Must the network receive a new functional layer? Not only addressing network management and service delivery, but trustworthiness and integrity as a separate functionality, as illustrated in the accompanying figure.

Conclusion

The step from 5G to 6G is not small, in fact, it's huge; as the vision of 6G enabling personal mobile robotics and XR requires far more than an "update." Maybe even a new layer in network operations should be included, as trustworthiness will be a process not only for designing systems but must also be guaranteed during services. This is a grand challenge for electrical and computer engineering. 

References

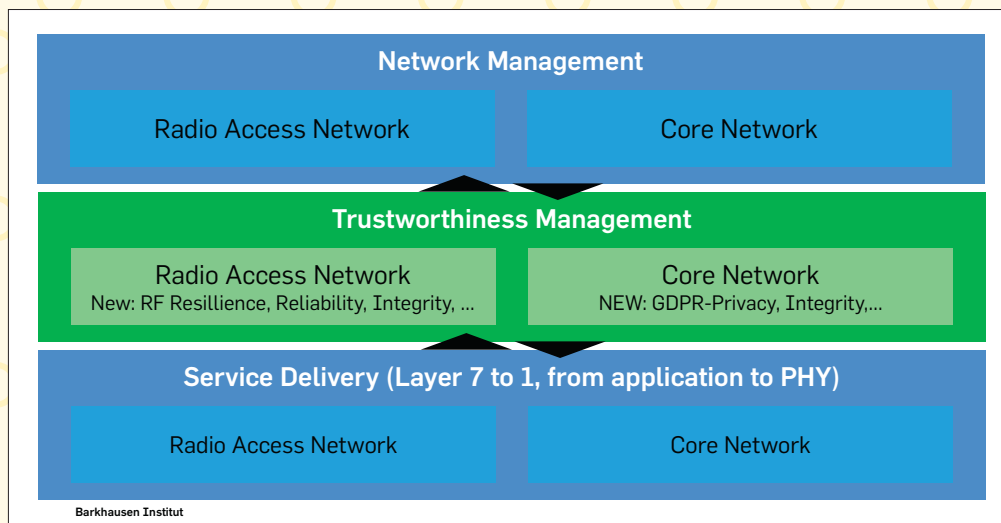
1. Ahlswede, R., Csiszar, I. On oblivious transfer capacity. *Information Theory, Combinatorics, and Search*. LNCS 7777 (2013). Springer.
2. Ahlswede, R. and Dueck, G. Identification via channels. *IEEE Trans. Info. Theory* 35, 1 (Jan. 1989), 15–29.
3. Asmussen, N. et al. M3: A hardware/operating-system co-design to tame heterogeneous manycores. *ASPLOS 2016*.
4. Fettweis, G. The tactile Internet: Applications and challenges. *IEEE Vehicular Tech. Mag.* 9, 1 (Mar. 2014), 64–70.
5. Fettweis, G. Beyond 5G: What could it be? Presentation at IEEE 5G Summit Dresden, 2019; <http://www.5gsummit.org/dresden-2019/>
6. Fettweis, G. Boche, H. 6G: The personal tactile Internet—Open questions for information theory. *IEEE BITS the Info. Theory Mag.* DOI: 10.1109/MBITS.2021.3118662
7. Fettweis, G. et al. Joint Communications and Sensing. ITG Position paper; <https://bit.ly/3EGwTcg/>
8. Höbner, T. et al. Mission availability for wireless URLLC. In *Proceedings of IEEE Globecom*, 2019; DOI: 10.1109/GLOBECOM38437.2019.9013362.
9. Köpsell, S. et al. Open-RAN risk analysis. BSI Study (in German); <https://bit.ly/3zbd6kj>
10. Pereg, U. et al. Classical State Masking over Quantum Channels. Sept. 2021; arXiv:2109.12647.
11. Vaikuntanathan, V. Homomorphic Encryption References; <https://people.csail.mit.edu/vinodv/FHE/FHE-refs.html>
12. Vilanova, L. et al. Caladan: A distributed meta-OS for data center disaggregation. In *Proceedings of the 10th Workshop on Systems for Post-Moore Architectures*, 2020.

Gerhard P. Fettweis is Scientific Director and CEO of the Barkhausen Institute, and Vodafone Chair Professor at TU Dresden, Germany. He coordinates the 5G++Lab Germany, and is PI at centers: 6G-life, CeTI, 5G++Lab Germany, and EKfZ.

Holger Boche is a professor and Chair at TU Munich. He coordinates the 6G Hub "6G-life" in Germany, and is PI at centers: 6G-life, CAsA, and MCQST.

The authors thank their funding agencies and companies, particularly the centers listed in their affiliations.

Copyright held by authors/owners. Publication rights licensed to ACM.



Projected need for introducing a new **Trustworthiness Management Layer (green)**.

Internet of Production—Entering Phase Two of Industry 4.0

BY GERTRUDE KAPPEL, CHRISTIAN BRECHER, MATTHIAS BROCKMANN, AND ISTVÁN KOREN

MAKING A HIGH-QUALITY gear cannot be learned simply from an Internet search. You may find guidelines, papers, rules, lectures, and videos. However, applying this general knowledge to a specific production process and dealing with uncertainties and disruptions requires special know-how, most of which resides in people's heads and networks and is acquired to a large extent through "learning by doing."

Over 10 years ago, the vision of Industry 4.0⁵ was announced at the Hannover Fair 2011 as part of the German/European High-Tech Strategy and adopted internationally by the Japanese Industrial Value Chain Initiative, the Advanced Manufacturing Initiative in the U.S., the Chinese Made in China 2025 strategy, the South Korean Manufacturing 3.0, and the U.K.'s High-Value Manufacturing Catapult research center. This "fourth industrial revolution" follows the earlier stages of mechanization (steam engine), mass production (assembly lines), and IT-based electronic automation. Core elements of Industry 4.0 include a reference architecture (RAMI 4.0)

for networking increasingly autonomous IoT devices, and cyber-physical production systems (CPPS) where simulations ("Digital Twins") help monitor, predict, and control physical production systems. Moreover, personalized and context-specific "assistants" and related organization forms should enable "new work" settings.

Yet, the vision of Industry 4.0 is far from being realized. In the research cluster "Integrative Production Technology for High-Wage Countries" (2006–2018),² the mapping of interdisciplinary models together with advanced mathematical techniques for model-order reduction not only resulted in much

more realistic models, but also accelerated their simulation by up to five orders of magnitude. Meanwhile, data-driven analytics and machine learning (ML) entered the production landscape. The Excellence Cluster "Internet of Production" (IoP, 2019–2025)⁴ aims at a new level of digital collaboration with data, models, and knowledge in production. The core idea is establishing a "worldwide lab" (WWL) for cross-domain learning, breaking down current data silos (see Figure 1). A WWL combines data analytics, domain-specific models, and expert knowledge, and sharing of the gained know-how for a large industrial domain.

To create real added value from the gigantic amounts of data that exist in all areas of production, a complete Digital Twin seems completely unrealistic due to the size of high-resolution databases, network overload, security, and data sovereignty concerns. Instead, the IoP approach circumscribes a near real-time digital representation of the WWL by a large collection of inter-related Digital Shadows (DS) as depicted in Figure 2. As task- and context-dependent, purpose-driven, aggregated, and persistent datasets, DS encompass a complex reality from multiple perspectives in a more compact fashion and with better performance than

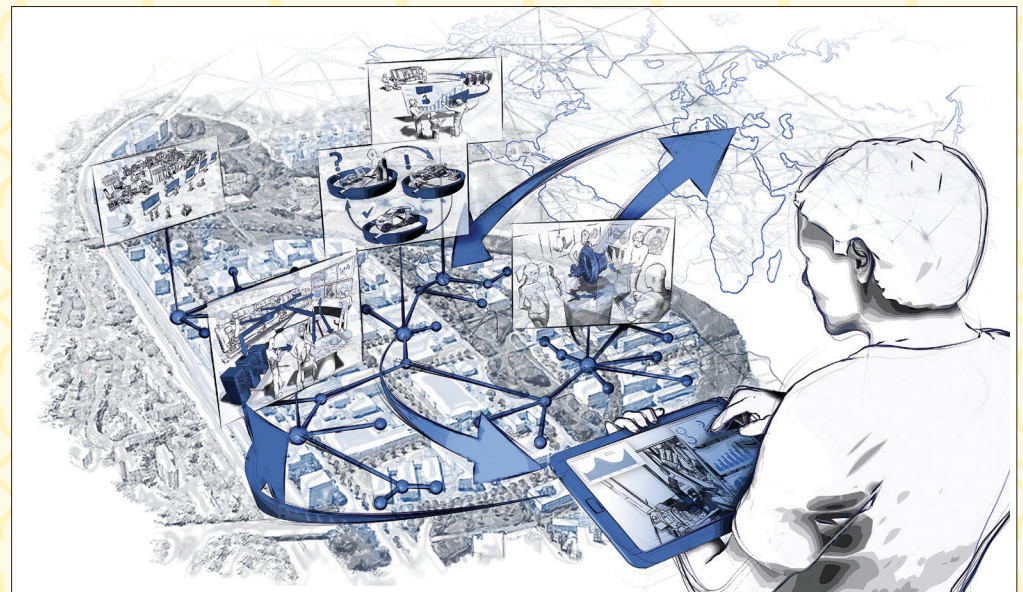


Figure 1. The worldwide lab of production is a breeding ground for innovation and cross-domain learning. Source: Martin Riedel.

a fully integrated Digital Twin. From a production management perspective, DS can be interpreted as mediators between the vast amounts of heterogeneous data and detailed production engineering models with the needed granularity level.⁸

DS research involves several CS disciplines.¹ From a data management and AI perspective, a DS is an extended database view, comprising a defining reduced mathematical or ML model (view definition query) and its partial stored answer (materialization).⁶ From a model-driven software engineering perspective, traceability/provenance and cross-view interrelationships between different DS need to be managed. To complete the infrastructure,⁷ process mining creates and interprets DS on large-scale event sequences. HCI considers user interfaces to DS, and communication engineering studies DS as wrapped objects of transportation, protection, and storage in


a network.

In the IoP cluster, more than 30 IoP institutes from different disciplines, such as computer science, engineering, material science, economics, and social sciences, as well as more than 50 industrial partners contribute competence in an iterative exploration and validation process, studying over a dozen interdisciplinary use cases. Application examples in aircraft turbine production, steel rolling, plastics and textile engineering, as well as electric car life-cycle engineering improve traditional KPIs such as quality, cost or efficiency, but also address important sustainability goals such as waste avoidance, energy saving, and CO₂ reduction. Initial industrial uptake is happening by SMEs and start-ups, but also in the context of industry platforms such as the Volkswagen “Industrial Cloud,” the BMW “Open Manufacturing Platform,” or the

The vision of Industry 4.0 is far from being realized.

alliance-driven “ADA-MOS” Industrial IoT platform. They enable new engineering solutions, for example, a new form of condition monitoring for a wide field of engineering applications.

In the future, the value of real production data—labeled by means of a suitable standard—will increase drastically. To create and fairly capture this added value, the world of production in academia and industry must advance toward a trustworthy WWL. Beyond integrating engineering and CS/AI research, this requires more socio-economic research in considering DS as social objects of human work, and as the units of trade in business models. In summary, we understand the concept of Digital

Shadows as an important novel computer science abstraction for representing and sharing CPPS knowledge in future production environments. 

References

1. Brauner, P. et al. A computer science perspective on digital transformation in production. *ACM Trans. Internet Things* 3, 2 (2022); <https://doi.org/10.1145/3502265>
2. Brecher, C. (Ed.) *Integrative Production Technology for High-Wage Countries*. Springer 2012.
3. Brecher, C., Obdenbusch, M., Herfs, W. Towards optimized machine operations by cloud integrated condition estimation. *Machine Learning for Cyber Physical Systems* (2016), 23–31; https://doi.org/10.1007/978-3-662-48838-6_4
4. Internet of Production Cluster of Excellence; <https://www.iop.rwth-aachen.de/cms/-gpfz/Produktionstechnik/?lid=1>
5. Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., and Hoffmann, M. Industry 4.0. *Business & Info. Syst. Engineering* 6, 3 (2014), 239–242.
6. Liebenberg, M., Jarke, M. Information engineering with digital shadows: Concept and case studies. In *Proceedings of CAISE (Grenoble, France, 2020)*, Springer LNCS 12127, 70–84.
7. Pennekamp, J. et al. Towards an infrastructure enabling the Internet of Production. In *Proceedings of IEEE Intl. Conf. Industrial CPS (Taipei, Taiwan, 2019)*, 31–37.
8. Riesener, M., Schuh, G., Dölle, C., Tönnies, C. The Digital Shadow as enabler for data analytics in product life cycle management. *Procedia CIRP* 80 (2019) 729–734; <https://doi.org/10.1016/j.procir.2019.01.083>

Gertrude Kappel is a professor for Business Informatics at Vienna University of Technology, Austria.

Christian Brecher is a professor at the Chair for Machine Tools at the Laboratory for Machine Tools and Production Engineering (WZL) and is the Spokesperson of the Cluster of Excellence “Internet of Production” at RWTH Aachen University, Germany.

Matthias Brockmann was managing director of the Cluster of Excellence “Internet of Production” at RWTH Aachen University, Germany.

István Koren is deputy area coordinator of the Cluster of Excellence “Internet of Production” at RWTH Aachen University, Germany.

Copyright held by authors/owners. Publication rights licensed to ACM.

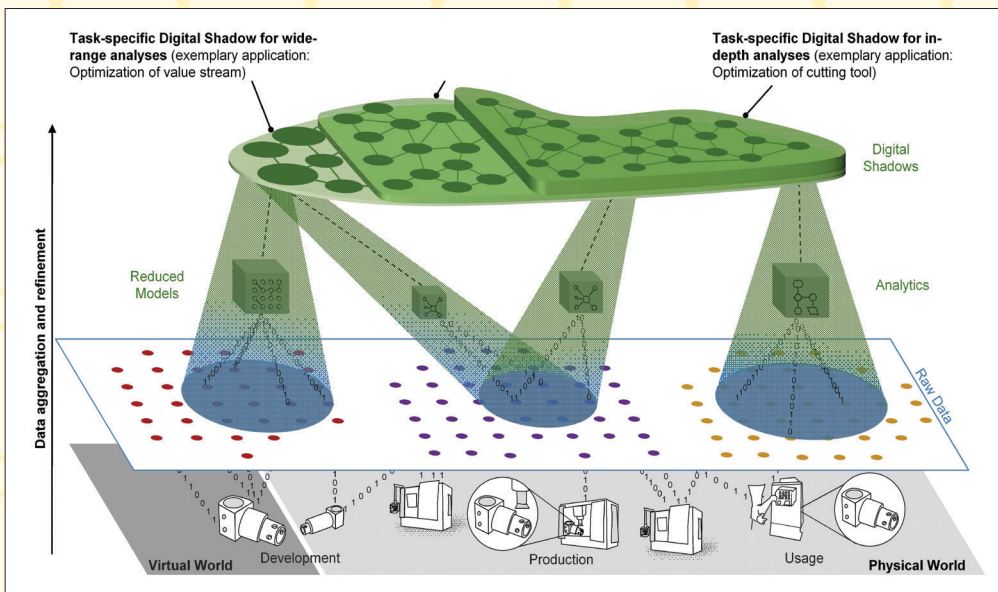


Figure 2. Concept of the Digital Shadow in production. Source: Cluster of Excellence “Internet of Production.”

Privacy-Preserving AI for Future Networks

BY DIEGO PERINO, KLEOMENIS KATEVAS, ANDRA LUTU, EDUARD MARIN, AND NICOLAS KOURTELLIS

TELCO NETWORKS AND systems evolved over the years to deal with novel services. Today, they are highly complex, distributed ecosystems composed of very diverse sub-environments (see Figure 1). They include myriad types of devices, connectivity means, protocols, and infrastructures often managed by different teams with varying expertise and tools, or even different companies.

Traditional network management solutions (for example, network over-provisioning, rule-based systems, reactive approaches) are reaching their limits in dealing with this complex ecosystem. Novel solutions are required to guarantee strict service requirements and effective resource management, especially in cases where entities have a partial view of the system.

AI to the rescue? In the last decade, we have witnessed a growing interest within the networking research community toward artificial intelligence (AI),^a whose techniques have been applied to a wide range of use cases: network optimization, routing, scheduling algorithms, resource and fault manage-

^a We refer to artificial intelligence as the set of tools falling in the category of machine learning and deep learning.

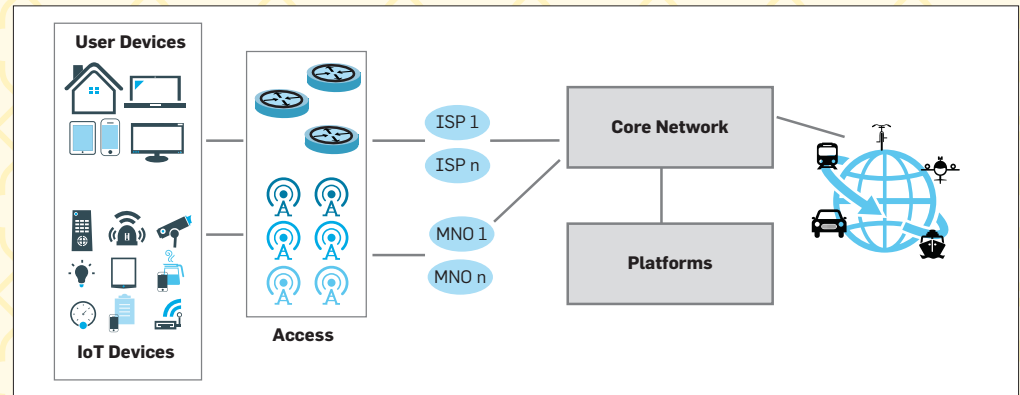


Figure 1. High-level view of the complexity of telcos' networks and systems with a large variety of devices, connectivity means, protocols, and infrastructures.

ment mechanisms, Quality of Service (QoS) and Quality of Experience (QoE) management, network security and many other tasks.¹ On the industrial side, AI is slowly complementing traditional networking approaches worldwide: Small Medium Enterprises (SMEs) and start-ups are developing AI solutions to deal with specific use cases, traditional networking vendors are evolving their products to support AI tools, and major cloud/software providers are adapting AI tools to be used in the networking domain. This is also the case for telephone service providers (telcos), which are exploring the application of AI algorithms through internal research and innovation (R&I) projects. For instance, our group is working on AI-based approaches in many use-cases focusing on realistic environments and applications (for example, Kattadige et al.² and Perino

et al.³), partially in collaboration with other European partners in the context of European R&I actions.^b

What about my privacy?

AI models and tools are potentially vulnerable, and their usage introduces new attack vectors for telco environments. For instance, membership and

property inference or data reconstruction attacks, and adversarial learning, can reveal different aspects of the data used (for example, which specific users' data were used for model training), the values of their data attributes, and even user patterns such as their mobility or browsing behavior. Therefore, the use of AI techniques can impact user privacy more strongly than traditional data analysis methods since AI models can distill information from multiple data sources and infer rich patterns regarding

^b CONCORDIA (<https://www.concordia-h2020.eu>), DAE-MON (<https://h2020daemon.eu>), ACCORDION (<https://www.accordion-project.eu>), CHARITY (<https://www.charity-project.eu/en>), SPATIAL (<https://spatial-h2020.eu>)

AI models and tools are potentially vulnerable, and their usage introduces new attack vectors for telco environments.

their data owners. Further exacerbating the privacy problem, data and model poisoning attacks can even manipulate AI models to take adversarial decisions for targeted users. Thus, recently enforced data privacy regulations (first in EU with GDPR and e-Privacy, since 2020 in U.S. with CCPA, and elsewhere in the world) attempt to mitigate these risks with specific guidelines to data operators/processors using AI methods.

To address this challenge, to follow regulations, and to build systems able to guarantee privacy by design, the R&I community is investigating the use of privacy-preserving AI (PPAI) methods, including techniques such as federated learning (FL), differential privacy, policy-based AI, and trusted execution environments (TEEs).


Our research group is building on these techniques with focus on the

complex telco ecosystems. Indeed, the capabilities at the edge and network devices (for example, IoT, phones, routers, antennas) can be used to perform the computation of AI models in a distributed hierarchical fashion without the need of sharing the data, and thus limiting the risk of private information leakage. This could also be done following an “as a service” approach to facilitate AI model building in a collaborative fashion between companies and mitigating attacks against FL model building using TEEs, as shown in our recent works^{3,4} and Figure 2). Furthermore, there are major trade-offs between privacy and utility of PPAI,⁶ and especially when introducing hierarchies in the FL process.

What's Next?

R&I is still needed to create autonomous, intelligent,

and yet fully privacy-preserving and secure telco “networks.” Creating customized AI-based approaches that tackle the specific challenges of network-related problems require more work. Particular attention should be devoted to the trade-off between creating complex tools that guarantee the required level of performance and robustness, and tools that network engineers trust and use. Interestingly, for many scenarios, it is still unclear whether AI is superior to traditional approaches, and how to seamlessly complement traditional methods with AI.

Further R&I is also required to design PPAI mechanisms suitable for these environments, including hierarchical FL, on-device training optimizations, or adaptive mechanisms able to deal with heterogeneous computing environments. 

References

1. Casas, P. Two decades of AI4NETS—AI/ML for data networks: Challenges and research directions. In *Proceedings of IEEE/IFIP NOMS*, 2020.
2. Kattadige, C., Raman, A., Thilakarathna, K., Lutu, A., Perino, D. 360NorVic: 360-degree video classification from mobile encrypted video traffic. In *Proceedings of ACM NOSSDAV*, 2021.
3. Kourtellis, N., Katevas, K. and Perino, D. FLaaS: Federated learning as a service. In *Proceedings of the ACM CoNext Distributed ML*, 2020.
4. Mo, F., Haddadi, H., Katevas, K., Marin, E., Perino, D. and Kourtellis, N. Privacy-preserving federated learning with TEEs. In *Proceedings of ACM MobiSys*, 2021 (best paper).
5. Perino, D., Yang, X., Serra, J., Lutu, A. and Leontiadis, I. Experience: Advanced network operations in (Un)-connected remote communities. In *Proceedings of ACM MobiCom*, 2020.
6. Zhao, B.Z.H., Kaafar, M.A., Kourtellis, N. Not one but many trade-offs: Privacy vs. utility in differentially private machine learning. In *Proceedings of ACM CCSW*, 2020.

Diego Perino is Director of Research at Telefonica Research, Spain.

Kleomenis Katevas is a Research Scientist at Telefonica Research, Spain.

Andra Lutu is Senior Research Scientist at Telefonica Research, Spain.

Eduard Marin is a Research Scientist at Telefonica Research, Spain.

Nicolas Kourtellis is Senior Research Scientist at Telefonica Research, Spain.

Copyright held by authors/owners. Publication rights licensed to ACM.

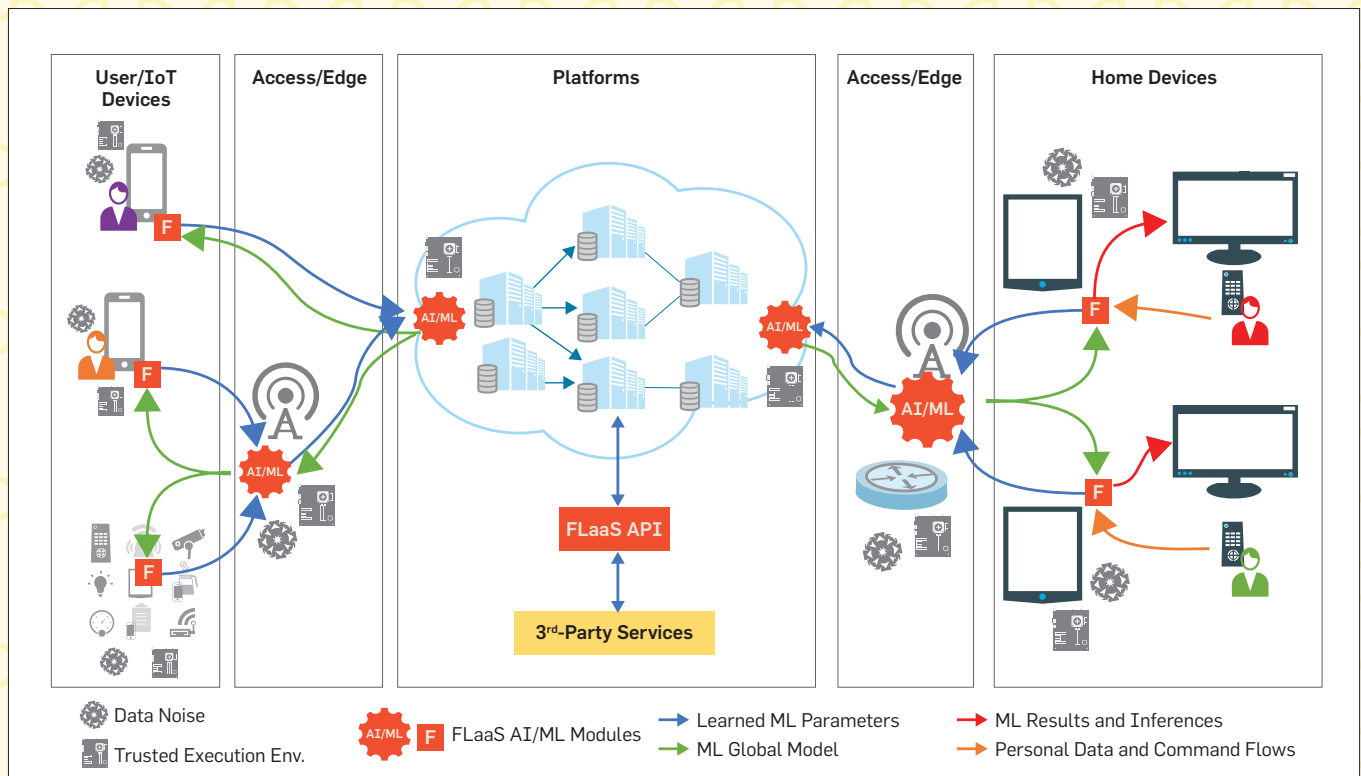


Figure 2. Example of Privacy Preserving Federated Learning in a telco network. Federated Learning is provided in an “as a service” approach with potential attacks mitigated by the usage of TEEs, data noise, and hierarchical model building.

Partnership on AI, Data, and Robotics

BY EDWARD CURRY, FREDRIK HEINTZ, MORTEN IRGENS, ARNOLD W.M. SMEULDERS, AND STEFANO STRAMIGIOLI

IN HER 2020 State of the Union Address,^a President of the European Commission von der Leyen called for Europe to lead the way on digital in the areas of data and artificial intelligence (AI). Artificial intelligence, data and robotics (ADR) present an opportunity and a challenge for Europe, a chance to improve the competitiveness of the European public and private sectors, and a challenge to translate Europe's core AI, data, and robotics strengths into a global market advantage (see Figure 1).

Working together, the Big Data Value Association (BDVA), the Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE), the European Laboratory for Learning and Intelligent Systems (ELLIS), the European Association for Artificial Intelligence (EurAI), and the European Robotics Association (euRobotics) have founded the AI, Data and Robotics Association (Adra) in order to establish an effective European Partnership on AI, Data and Robotics with the European Commission. The objective is to strengthen European competitiveness, societal well-being, and environ-



mental sustainability. The vision is to lead the world in researching, developing, and deploying value-driven trustworthy AI, data, and robotics based on fundamental European rights, principles, and values.

The partnership is unique in bringing together AI, data, and robotics disciplines within a single initiative with a budget of €2.6 billion (US \$3 billion). Adra represents the private side of the partnership and developed the Strategic Research, Innovation and Deployment Agenda (SRIDA)⁴ to guide the partnership's work.

To achieve this vision, a

will only accept AI, data, and robotics products and technologies when they both trust them and see their value.

The AI, Data and Robotics Innovation Ecosystem Enablers represent activities in the ecosystem that underlie innovation across sectors and from research to deployment. To meet the goal, a substantial development in skills and knowledge is needed in European industry. For AI, data, and robotics to develop further, large volumes of cross-sectorial, unbiased, high-quality, and trustworthy data must become available. Data spaces, platforms, and marketplaces are enablers, the key to unleashing the potential of such data.¹ Experimentation and sandboxes are critical for ADR-based innovation because of the need to deploy in complex physical and digital environments.

Cross-Sectorial AI, Data and Robotics Technology Enablers represent the core technical competencies essential for developing successful AI, data, and robotics systems, services, and products. The sensing and perception, and knowledge and learning technology enablers create the data and knowledge on which decisions are made. These are used by the reasoning and decision-making technologies to deliver; edge

a https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_20_1655

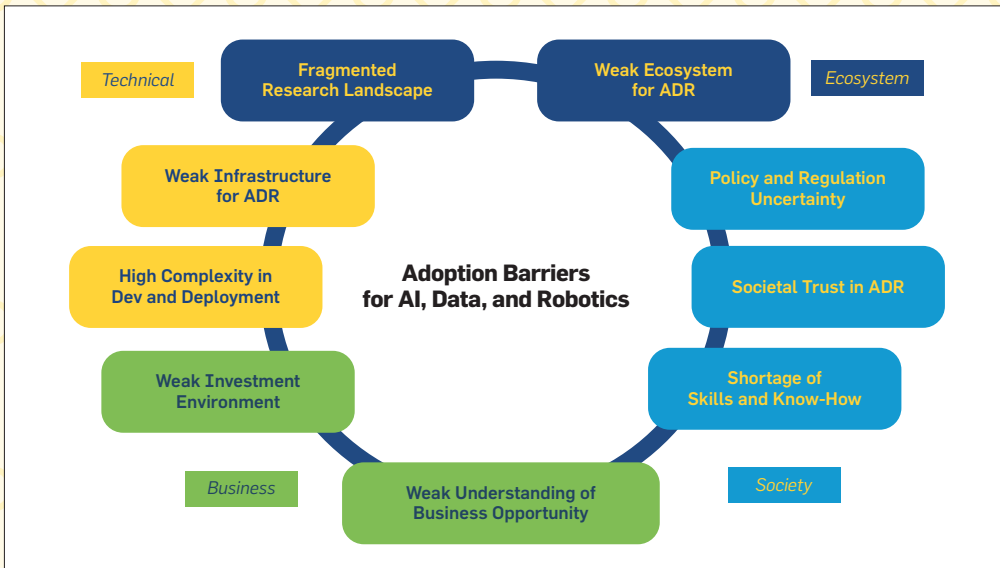


Figure 1. Challenges for adoption of AI, data, and robotics in Europe.

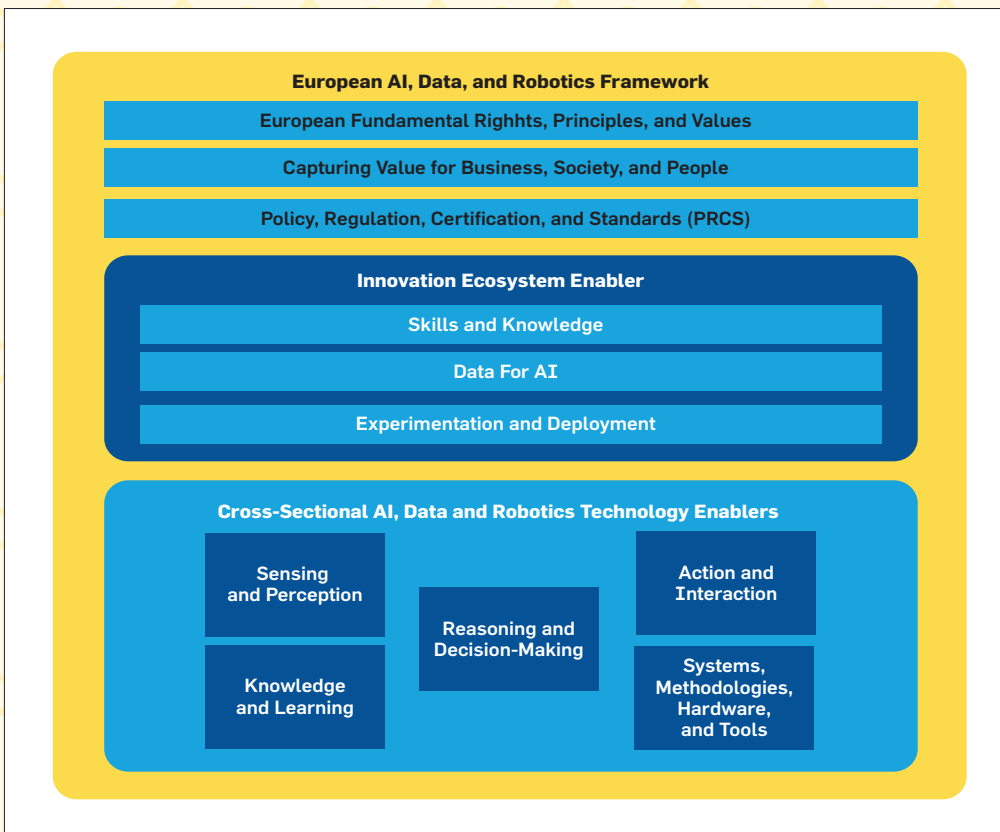


Figure 2. European AI, data, and robotics framework and enablers.


and cloud-based decision making, planning, search and optimization in systems, and the multilayered decision-making necessary for AI, data and robotics systems operating in complex environments. Action and interaction covers the challenges of human

interaction, machine-to-machine inter-operation, and machine interaction with the human environment, which can be even physical in the case of robotics applications. Finally, systems, methodologies, hardware, and tools provide methods

that enable the construction and configuration of systems by integrating technologies into systems and ensuring core systems properties, such as safety, robustness, dependability, and trustworthiness, are met. Each technical enabler overlaps with the

other, and there are no clear boundaries. Indeed, exciting advances are most often made in the intersections between these five areas and the system-level synergies that emerge from their interconnections.

Conclusion

The partnership on AI, data, and robotics will mobilize the ADR ecosystem in Europe to provide strong leadership in these areas, both in science, innovation, and deployment. It will create dialogues that address fundamental issues around deployment and citizen trust in AI. It will enable a rich AI, data, and robotics innovation ecosystem built on Europe’s many strong components, from its strong academic excellence, strong skills pipeline, and global companies to its innovation-driving regulation and standards coupled to best practice. 

References

1. European Commission. Communication: A European strategy for data (2020).
2. European Commission. White Paper on Artificial Intelligence—A European approach to excellence and trust (2020).
3. High-Level Expert Group on Artificial Intelligence. *Ethics Guidelines for Trustworthy AI* (2019).
4. Zillner, S. et al. *Strategic Research, Innovation and Deployment Agenda—AI, Data and Robotics Partnership*. Third Release, 2020, BDVA, euRobotics, ELLIS, EurAI and CLAIRe.


Edward Curry is a professor at the Insight SFI Research Centre for Data Analytics, NUI Galway, Ireland, and Vice-President at BDVA.

Fredrik Heintz is a professor in the Department of Computer Science at Linköping University, Sweden, and director of EU collaborations at EurAI.

Morten Irgens is Dean at Kristiania University College, Oslo, Norway, and Director of Innovation at CLAIRe.

Arnold W.M. Smeulders is a professor at the University of Amsterdam, the Netherlands, and an envoy of ELLIS.

Stefano Stramigioli is a professor of Advanced Robotics at the University of Twente, the Netherlands, and Vice-President Research at euRobotics.

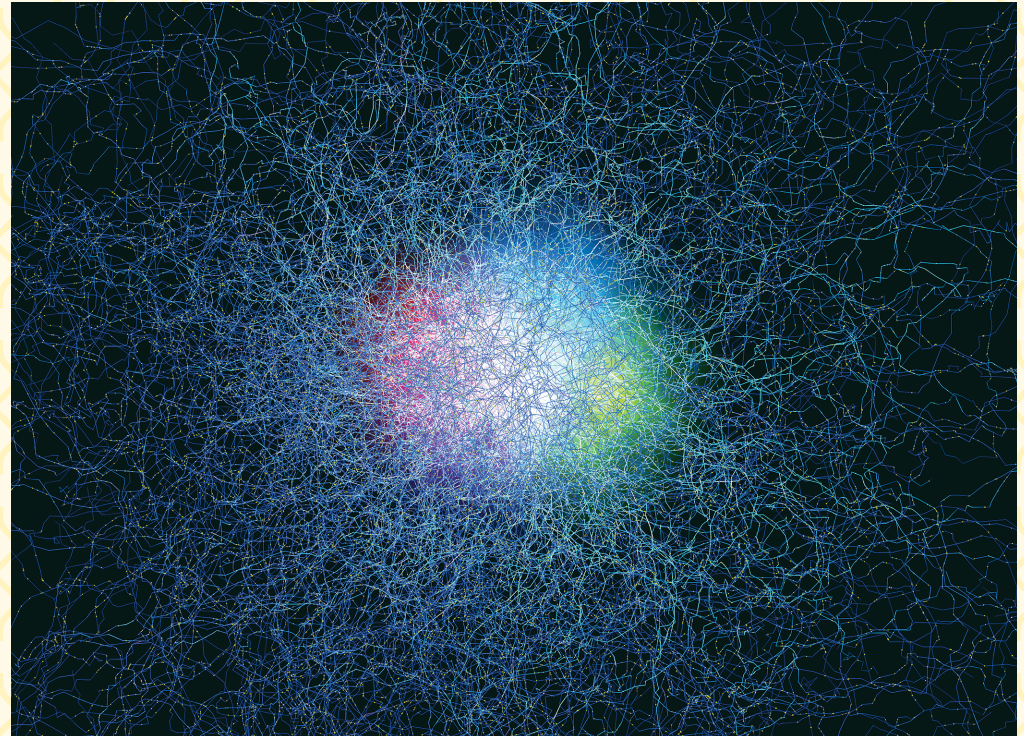
 This work is licensed under a Creative Commons Attribution 4.0 International License. <http://creativecommons.org/licenses/by/4.0/>

Toward a Broad AI

BY SEPP HOCHREITER

DESPITE BIG SUCCESSES in artificial intelligence (AI) and deep learning, there have been critical assessments made to current deep learning methods.⁸ Deep learning is data hungry, has limited knowledge transfer capabilities, does not quickly adapt to changing tasks or distributions, and insufficiently incorporates world or prior knowledge.^{1,3,8,14} While deep learning excels in natural language processing and vision benchmarks, it often underperforms at real-world applications. Deep learning models were shown to fail at new data, new applications, deployments in the wild, and stress tests.^{4,5,7,13,15} Therefore, practitioners harbor doubt over these models and hesitate to employ them in real-world application.

Current AI research has tried to overcome the criticisms and limitations of deep learning. AI research and machine learning in



particular aims at a new level of AI—a “broad AI”—with considerably enhanced and broader capabilities for skill acquisition and problem solving.³ We contrast “broad AI” to “narrow AI,” which are the AI systems currently applied. A broad AI considerably surpasses a narrow AI in the following essential properties:

A broad AI is a sophisticated and adaptive system, which successfully performs any cognitive task by virtue of its sensory perception, previous experience, and learned skills.

knowledge transfer and interaction, adaptability and robustness, abstraction and advanced reasoning, and efficiency (as illustrated in the accompanying figure). A broad AI is a sophisticated and adaptive system, which successfully performs any cognitive task by virtue of its sensory perception, previous experience, and learned skills.

To improve adaptability and robustness, a broad AI utilizes *few-shot learning*, *self-supervised learning* with contrastive learning, and processes sensory inputs using context and memory. Few-shot learning trains models with a small amount of data using prior knowledge or previous experience. Few-shot learning has a plethora of real-world applications, for example,

when learned models must quickly adapt to new situations, for new customers, new products, new processes, new workflows, or new sensory inputs.

With the advent of large corpora of unlabeled data in vision and language, self-supervised learning based on *contrastive learning* became very popular. Either views of images are contrasted with views of other images or text descriptions of images are contrasted with text descriptions of other images. Contrastive Language-Image Pre-training (CLIP)¹⁰ yielded very impressive results at zero-shot transfer learning. The CLIP model has the potential to become one of the most important foundation models.² A model with high zero-shot transfer learning

performance is highly adaptive and very robustness, thus is supposed to perform well when deployed in real-world applications and will be trusted by practitioners.

A broad AI should process the input by using context and previous experiences. Conceptual short-term memory⁹ is a notion in cognitive science, which states that humans, when perceiving a stimulus, immediately associate it with information stored in the long-term memory. Like humans, machine learning and AI methods should “activate a large amount of potentially pertinent information,”⁹ which is stored in episodic or long-term memories. Very promising are *Modern Hopfield networks*,^{11,12,16} which reveal the covariance structures in the data, thereby making deep learning more robust. If features co-occur in the data, then modern Hopfield networks amplify this co-occurrence in samples that are retrieved. Modern Hopfield networks are a remedy for learning methods that suffer from the “explaining

away” problem. Explaining away is the confirmation of one cause of an observed event that prevents the method from finding alternative causes. Explaining away is one reason for short-cut learning⁵ and the Clever Hans phenomenon.⁷ Modern Hopfield networks avoid explaining away via the enriched covariance structure.


Graph neural networks (GNNs) are a very promising research direction as they operate on graph structures, where nodes and edges are associated with labels and characteristics. GNNs are the predominant models of neural-symbolic computing.⁶ They describe the properties of molecules, simulate social networks, or predict future states in physical and engineering applications with particle-particle interactions.

Europe’s Opportunity for a Broad AI

The most promising approach to a broad AI is a neuro-symbolic AI, that is, a *bilateral AI* that combines methods from symbolic and sub-symbolic AI. In contrast

Europe has strong research groups in both symbolic and sub-symbolic AI, therefore has the unprecedented opportunity to make a fundamental contribution to the next level of AI—a broad AI.

to other regions, Europe has strong research groups in both symbolic and sub-symbolic AI, therefore has the unprecedented opportunity to make a fundamental contribution to the next level of AI—a broad AI.

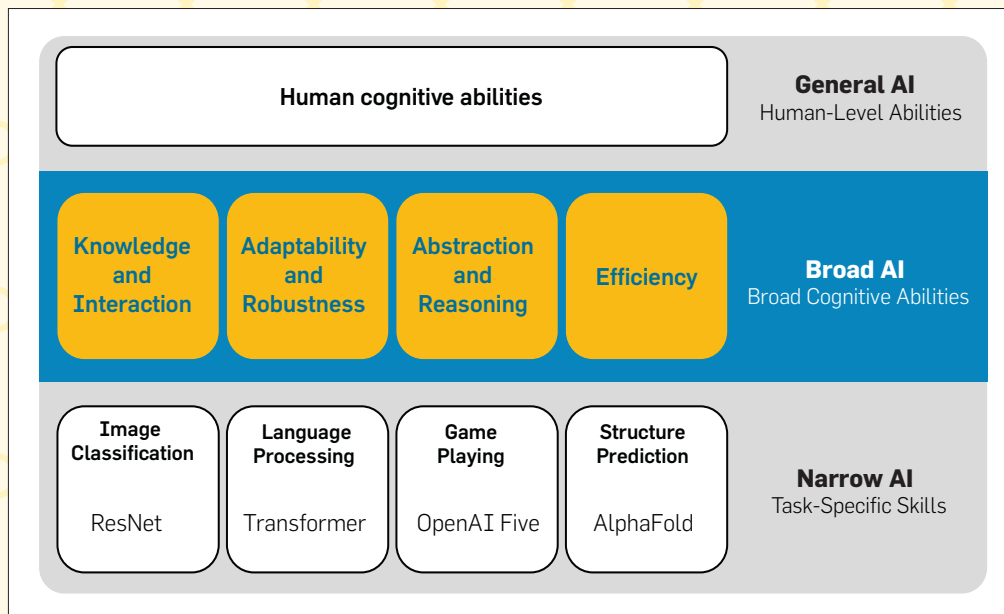
AI researchers should strive for a broad AI with considerably enhanced and broader capabilities for skill acquisition and problem solving by means of bilateral AI approaches that combine symbolic and sub-symbolic AI. 

References

- Bengio, Y., Lecun, Y., and Hinton, G. Turing lecture: Deep Learning for AI. *Commun. ACM* 64, 7 (July 2021), 58–65; doi:10.1145/3448250
- Bommasani, R. et al. On the Opportunities and Risks of Foundation Models (2021); *ArXiv:2108.07258*.
- Chollet, F. On the Measure of Intelligence (2019); *ArXiv:1911.01547*.
- D’Amour, A. et al. Underspecification Presents Challenges for Credibility in Modern Machine Learning (2020); *ArXiv: 011.03395*.
- Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R.S., Brendel, W., Bethge, M. and Wichmann, F.A. Shortcut Learning in Deep Neural Networks (2020); *ArXiv:2004.07780*.
- Lamb, L.C., Garcez, A., Gori, M., Prates, M., Avelar, P. and Vardi, M. Graph Neural Networks Meet Neural-Symbolic Computing: A Survey and Perspective (2020); *ArXiv:2003.00330*.
- Lapuschkin, S. et al. Unmasking clever Hans predictors and assessing what machines really learn. *Nature Communications* 10 (2019).
- Marcus, G. *Deep Learning: A Critical Appraisal*. (2018); *ArXiv:1801.00631*.
- Potter, M. Conceptual short term memory in perception and thought. *Frontiers in Psychology* 3 (2012), 113.
- Radford, A. et al. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th Intern. Conf. Machine Learning*. 2021.
- Ramsauer, H. et al. Hopfield Networks is All You Need (2020); *ArXiv: 2008.02217*.
- Ramsauer, H. et al. Hopfield networks is all you need. In *Proceedings of the 2021 Intern. Conf. Learning Representations*; <https://openreview.net/forum?id=tL89RnzIiCd>.
- Recht, B., Roelofs, R., Schmidt, L., and Shankar, V. (2019). Do ImageNet classifiers generalize to ImageNet? *Proceedings of the 36th Intern. Conf. Machine Learning* 97 (2019), 5389–5400.
- Sutton, R. *The Bitter Lesson*. 2019; <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>
- Taori, R., Dave, A., Shankar, V., Carlini, N., Recht, B., and Schmidt, L. Measuring robustness to natural Distribution shifts in image classification. In *Proceedings of the 33rd Conf. Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2020, 18583–18599
- Widrich, M. et al. Modern Hopfield networks and attention for immune repertoire classification. *Advances in Neural Information Processing Systems* 33 (2020); <https://bit.ly/3UPpI5y>.

Sepp Hochreiter is a professor at Johannes Kepler Universität Linz, Austria.

Copyright held by author/owner. Publication rights licensed to ACM.



Hierarchical model of cognitive abilities of AI systems.³

BY HARTMUT SCHMECK, ANTONELLO MONTI,
AND VEIT HAGENMEYER

Energy Informatics— Key Elements for Tomorrow's Energy System

THE INCREASINGLY VISIBLE effects of climate change necessitate a fundamental transformation of energy systems toward renewable sources. While the Fukushima event led to a particularly strong change in energy policies in Germany, resulting in the so-called *Energiewende*, or energy transition, the trend toward renewables is visible worldwide. Here, we outline how major challenges of the energy transition have led to a strong need for essential contributions from the computer science community to maintain stability and security of supply, particularly for the electric power grid. As a result, the new discipline of Energy Informatics has emerged which is addressing this highly interdisciplinary and dynamic field of research and development.

While there are numerous demanding aspects of the energy transition, there are several major problem areas which show why contributions from Energy Informatics are urgently needed:

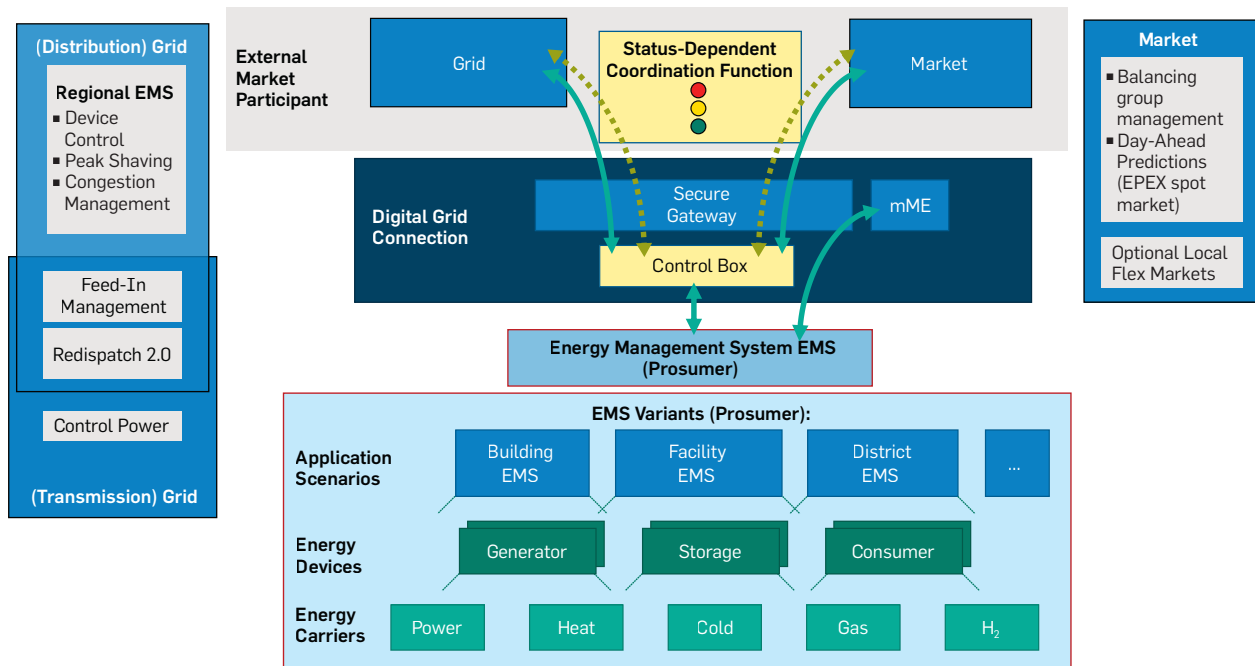
Volatility. In tomorrow's energy system, electric power will be provided mainly by photo-voltaic modules on rooftops and in larger field installations, and by wind power plants, onshore as well as offshore. Being weather-dependent, this energy supply is inherently volatile and only partially controllable. Therefore, the classic principle "supply follows demand" is no longer sufficient, rather we need to follow the supply with demand, which requires the demand side to become sufficiently flexible and adaptable. In this context, the optimal use of storage, for example batteries, will become very relevant.

Uncertainty. In addition to the increasingly intermittent character of renewable energy supply, predictions of energy supply and demand have become significantly more difficult. Beyond the weather-dependent uncertainty of supply, the behavior of consumers is changing—by using new applications such as heat pumps and electric vehicles, for example, and by using demand-side management in response to time-dependent pricing—and hence no longer necessarily corresponds to standard load profiles. Therefore, the transition and distribution system operators as well as the balancing responsible parties need more information on the actual behavior of end customers and their actual energy schedules, and they must respond more dynamically and more often, almost in real time, to observed deviations from expected energy schedules.

Decentralization. While in traditional power grids electricity was generated by a few large power plants, fed into the highest voltage level grid, transmitted over long distances, and distributed locally down to the low voltage connection points of the end customer, in tomorrow's power grid electricity will be



Figure 1. Simplified structural view on activities in a decentralized power grid.



generated at millions of low voltage locations where it will either be consumed directly or fed into the grid, which will have to collect and (re) distribute the energy bidirectionally. Local power in-feed can lead to congestions due to insufficient cable capacity resulting in voltage increase. The opposite effect can occur because of the emergence of large new loads from recharging the batteries of electric vehicles. Both effects are hardly visible to the distribution system operators (DSOs), even if they have remote information about the status of their substations. Another implication of the trend from centralized to decentralized power generation is the gradual disappearance of large power plants as the traditional providers of ancillary energy system services, again leading to an increasing demand for flexibility in decentralized energy schedules.

Modified dynamic characteristics of the power system. Traditionally power systems have been controlled as electromechanical systems. A grid dominated by power electronics is rather a software-defined system that is capable of dynamic response way faster than a traditional grid based on large power plants. Hence, we need faster control responses and smarter IT solutions.⁸

Obviously, the essential prerequisite for dealing with these major problems is the availability of information on the status of all active components of the power grid, which shows the need for digitalization. But while in the traditional power grid the transmission system operators (TSOs) were responsible for dealing with stability problems of frequency and voltage, in tomorrow's grid the DSOs and even the energy managers of facilities, buildings, and homes will have to respond locally to such quality problems, mainly with respect to voltage but to some extent even in response to frequency deviations as long as they are part of aggregated service providers. Hence, there is an inherent need for an energy information and control network with distributed system intelligence, and for a multitude of locations where adequate control decisions can be derived, mostly from locally available data.

Consequences. To cope with the inherent volatility, uncertainty, and decentralization of energy supply from renewable sources it is necessary to discover and exploit the flexibility of load schedules related to demand and supply. This can only be achieved by analysis of available data and by interaction with entities which are responsible for the operation of relevant

devices, including (smart) homes and buildings, industrial processes, and those related to the energy requirements of electric vehicles. For this, the computer science community is challenged to provide the necessary adequately designed methods and tools. In this way the tasks of informatics are extended from information and communication technologies to operational technologies and even to real-time control, which underlines requirements for a joint multidisciplinary effort of computer scientists, power engineers, and control engineers. Beyond technical problems they also must consider essential economic and legal issues in this highly regulated critical infrastructure. All of this is summarized in the term Energy Informatics. A simplified structural view on decentralized activities in the power grid is provided in Figure 1.

Essential Contributions of Energy Informatics

Due to strict space constraints, we can only highlight a few of the essential topics which have to be addressed by Energy Informatics and cannot include an adequate overview of the numerous European locations performing related research. Nevertheless, in addition to brief descriptions and examples of our

own research we include links and references to related publications, projects, and institutions which provide further links to relevant research activities.

Provide adequate information on current and historic status of the energy system. The prerequisite for providing information is the availability of sensors like smart meters which can measure the current status of energy-relevant devices, and of an infrastructure for delivering measured data values to authorized recipients. A common requirement is the capability to support bidirectional exchange of information to transmit information on dynamic, time-variant tariffs from energy suppliers, or to send control signals from active external market participants (like Demand-Side Managers or DSOs) to controllable local systems. There is a range of topics related to the design and utilization of such an infrastructure, like concerns with respect to security and privacy, or the appropriate choice of temporal and spatial granularity of data (see, for example, Kroener et al.⁶).


Data analytics for energy status data. As energy-related data will be available at high spatial and temporal resolution, intelligent methods have to be designed for utilizing this valuable source of information. In this context, funded by the German Research Foundation (DFG), an interdisciplinary Research Training Group^a spread over 11 research groups at Karlsruhe Institute of Technology is dedicated to informatics methods for various challenges in the life cycle of energy status data, consisting of *collection, analysis, deployment, and exploitation*.

Innovative and scalable architectures for data platforms. The increased level of decentralization calls for scalable data platforms that are active mostly at the edge. Classical centralized SCADA architectures are not able to cope with the huge amount of data that must be processed. Solutions in these directions are emerging, mostly from the world of the open-source communities, as with the SOGNO project^b at Linux Foundation Energy.


Modeling, (co-)simulation, and

prediction. The contributions of Energy Informatics crucially depend on adequate modeling and simulation of energy-relevant devices, systems, grids, and related processes, providing the ability to analyze and predict behavior under various conditions. This is particularly necessary for predicting potential congestions in the power grid, based on anticipated power flows in the grid and load profiles of relevant entities. New grid dynamics call for completely new modeling approaches for power systems, and corresponding software,⁷ which is also essential for the design of efficient wind energy systems, a major topic of the DTU Wind Energy Center.⁸ Beyond the separate simulation of individual devices or aggregated entities, their interaction must be analyzed using co-simulation. As an example, the potential contribution of intelligent buildings for reducing congestion in the distribution grid can be investigated by a multihome simulation in combination with a simulation of the distribution grid.⁵ An interesting comparison of co-simulation frameworks is provided in Steinbrink et al.¹⁰ A particularly relevant modeling task is the discovery of the potential flexibility in the load profiles of entities (see Barth et al.¹). Machine learning is essential for detecting degrees of freedom in the use of certain appliances by analyzing recorded load profile data streams (see Šikšnys⁹ and Förderer³).

Energy Management Systems (EMS). The anticipated energy information network with distributed system intelligence crucially depends on an efficient and effective management of energy at various levels of the grid. At the end-customer level a home, building, or facility EMS must provide effective visualization of the energy status and user-friendly interaction with humans to detect local preferences enabling adequate optimization of schedules for local supply and demand. The EMS has to serve as the digital connection point for the next layer (the DSO or the regional EMS), translating external requests for load changes into appropriate schedule changes for local devices. Such an EMS depends on the availability of data and methods for



The new discipline of Energy Informatics has emerged to address the strong need for essential contributions from the computer science community to maintain stability and security of supply, particularly for the electric power grid.



a <https://www.energystatusdata.kit.edu>

b <https://www.lfenergy.org/projects/sogno/>

c <https://windenergy.dtu.dk/english>

data analysis, simulation, prediction, and optimization. An example of such an EMS is the Organic Smart Home,^d which has been developed in a sequence of smart grid projects. A sample architecture for the interaction between an external market participant and a local charging station for an electric vehicle (EV) via a smart meter gateway and a local EMS is shown in Figure 2.

Cyber security issues. The inherently growing intrusion of information and communication technologies into energy systems inevitably leads to new vulnerabilities of this highly critical infrastructure. Therefore, security, safety, and data protection have emerged as essential topics for the design and operation of smart energy grids. A particular challenge is to reconcile the seemingly contradictory requirements for functionality, real-time capability, privacy protection, and robustness against attacks and disruptions. Distributed energy systems should not only have a secure IT infrastructure, but also be resilient since attacks cannot be completely avoided. Relevant research on security issues is available from the German Competence Center for Applied Security Technologies,^e the Queen’s University Belfast center for Secure Digital Systems,^f and the Norwegian Department of Information Security and Communication Technology.^g

d <https://organicsmarthome.fzi.de/>

e <https://zentrum.kastel.kit.edu/english/index.php>

f <https://pure.qub.ac.uk/en/organisations/secure-digital-systems-sds/>

g <https://www.ntnu.edu/iik/>

Regional Community on Energy Informatics

There is a growing research community on Energy Informatics in Europe which emerged mainly in the last decade. For example, the German Informatics Society established a special interest group on Energy Informatics which has strong ties with the newly formed ACM SIG Energy and its flagship conference, ACM e-Energy. Based on a joint German-Austrian-Swiss initiative, the DACH+ conference series on Energy Informatics^h evolved as an annual conference, moving cyclically through Germany (where it started in 2012), Austria, and Switzerland. Its objectives are to promote the research, development, and implementation of information and communication technologies in the energy domain and to foster the exchange between academia, industry, and service providers in the German-Austrian-Swiss region and its neighboring countries.

Large Infrastructures and Projects for Research on Energy Informatics

Energy Lab 2.0. The Energy Lab 2.0ⁱ at KIT is one of Europe’s largest research infrastructures for renewable energy and the energy transition. The intelligent networking of environmentally friendly energy generators and storage methods are investigated. In addition, energy systems of the future are simulated and tested based on real consumer data.⁴ A plant network links electrical, thermal, and chemical energy flows as well as new information and

h <https://energy-informatics2021.org/>

i <https://www.elab2.kit.edu/>

communication technologies. The research aims at improving the transport, distribution, storage, and use of electricity and thus supports the energy transition.

C/sells, SINTEG. In large-scale tests for the energy supply of the future and the digitalization of the energy sector in the German-funded *Smart Energy Showcase—Digital Agenda for the Energy Transition* (SINTEG)^j program, more than 300 companies, research institutions, and municipalities worked together from 2016 to 2020. They formed five model regions, in which they developed and tested solutions for the energy supply of the future. To take an example, the *C/sells*^k project focused on developing a smart, cellular energy system, supported by an infrastructure system that enabled data to be exchanged securely between different cells, and a coordination sequence through which grid operators could communicate and act quickly, and for the most part autonomously. It also involved platforms for trading regional energy and flexibilities, leading to new services and products.

European projects. There are a considerable number of research projects on Energy Informatics funded through the EU’s eighth Framework Programme Horizon 2020. One of the largest is the currently active *One Network for Europe* (OneNet) project.^l Its scope is to create a fully replicable and scalable architecture that enables the whole European electrical system to operate as a single system in which a variety of markets allows the universal participation of stakeholders at every level, from small consumers to large producers, regardless of their physical location (see Figure 3). Led by the Fraunhofer Institute for Applied Information Technology, the project brings together a consortium of over 70 partners, including grid operators, key IT players, leading research institutions, and two European associations for grid operators. OneNet aims at creating the conditions for a new generation of grid services able to fully exploit demand response, storage, and distributed generation while creating

j <http://www.sinteg.de/en/>

k <https://www.csells.net/en/>

l <https://onenet-project.eu/>

Figure 2. EV-charging controlled by an EMS with respect to directives from an active External Market Participant via a smart meter gateway complying also with consumer preferences (taken from Kroener et al.⁶)

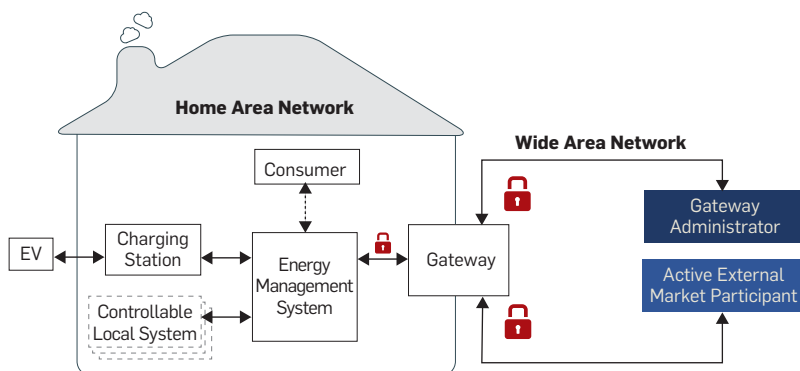
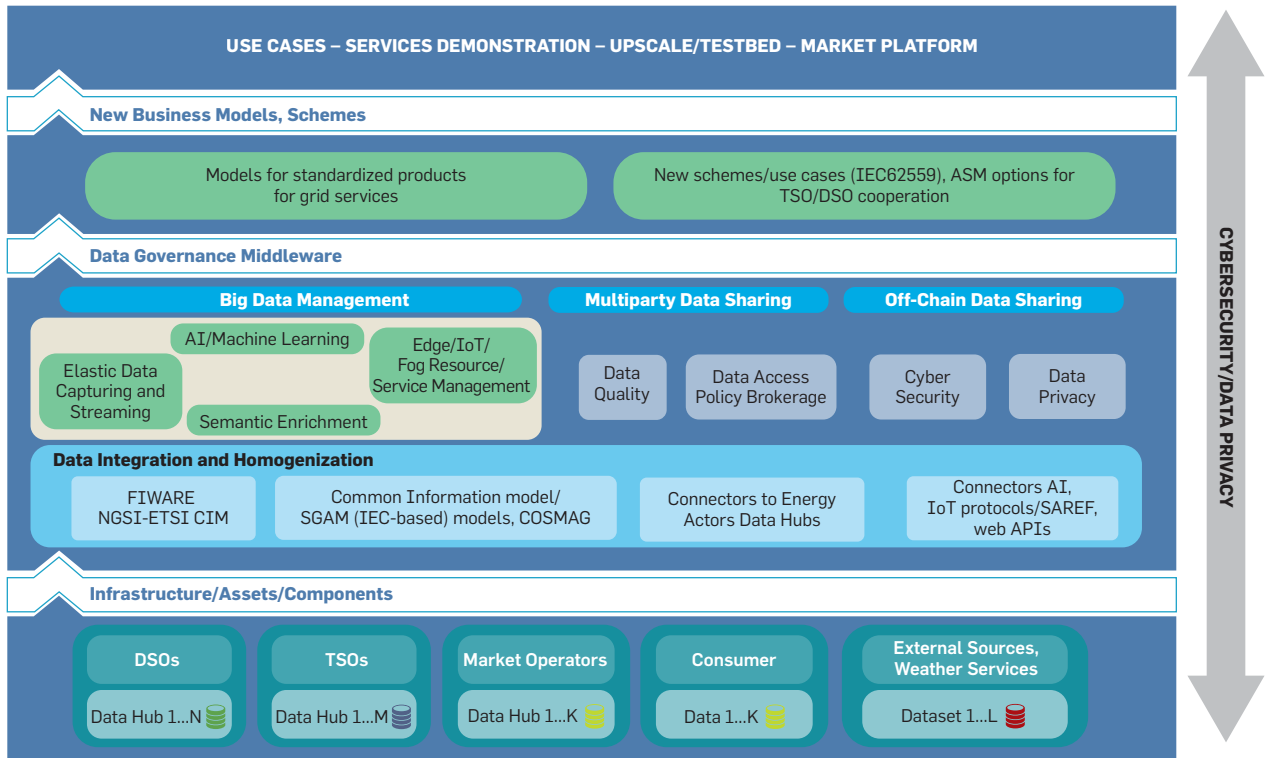



Figure 3. The vision of the OneNet project.



fair, transparent, and open conditions for the consumer.

Conclusion and Open Issues

The energy transition toward renewables is driving the energy system into an epochal transformation of its basic principles. The outlined combination of challenges can only be addressed through a real multidisciplinary approach and strong cooperation among the different energy sectors necessarily based on thorough digitalization. Energy Informatics can be seen as the main enabler that orchestrates all other elements and promises increased efficiency, new sources of load flexibility, and the necessary resilience to cope with inevitable local disturbances in a highly decentralized system. Tomorrow's energy system must be at least as reliable as the current one: new solutions are needed to go beyond the current practice, and in addition, lack of implementation competence in the field considering more complex system structures further increases the challenges of the transition. Beyond the gradual transformation of existing structure and management of energy systems there are also approaches to

design a fundamentally new type of power grid, inspired by the principles of the Internet of Data, dealing with energy packets and storage at every node of the grid (see De Din²), but it remains completely open to what extent this will lead to a viable concept. At least, intelligent utilization of ubiquitously available storage, mobile within EVs and stationary within buildings, will play a major role in tomorrow's energy system. 

References

1. Barth, L., Hagenmeyer, V., Ludwig, N., and Wagner, D. How much demand side flexibility do we need? Analyzing where to exploit flexibility in industrial processes. In *Proceedings of the 9th ACM Intern. Conf. Future Energy Systems*. June 2018, 43–62; <https://doi.org/10.1145/3208903.3208909>
2. De Din, E., Monti, A., Hagenmeyer, V., and Wehrle, K. A new solution for the energy packet-based dispatching using power/signal dual modulation. In *Proceedings of the 9th ACM Intern. Conf. Future Energy Systems*. June 2018, 361–365; <https://doi.org/10.1145/3208903.3208931>
3. Förderer, K., Ahrens, M., Bao, K., Mauser, I., and Schmeck, H. Towards the modeling of flexibility using artificial neural networks in energy management and smart grids. In *Proceedings of the 9th ACM Intern. Conf. Future Energy Systems*. June 2018, 85–90; <https://doi.org/10.1145/3208903.3208915>
4. Hagenmeyer, V., et al. Information and communication technology in energy lab 2.0: Smart energies system simulation and control center with an open-street-map-based power flow simulation example. *Energy Technology* 4, 1 (Jan. 016) 145–162; <https://doi.org/10.1002/ente.201500304>
5. Kochannek, S., Mauser, I., Phipps, K., and Schmeck, H. Hardware-in-the-loop co-simulation of a smart

- building in a low-voltage distribution grid. In *Proceedings IEEE PES Innovative Smart Grid Technologies Conf. Europe*, 2018, 1–6; <https://doi.org/10.1109/ISGTEurope.2018.857174>
6. Kroener, N., Förderer, K., Lösch, M., and Schmeck, H.: State-of-the-art integration of decentralized energy management systems into the German smart meter gateway infrastructure. *Applied Science* 10, 11 (2020), 3665; <https://doi.org/10.3390/app10113665>
 7. Mirz, M., Vogel, S., Reinke, G., and Monti, A. DPsim—A dynamic phasor real-time simulator for power systems. *SoftwareX* 10, (2019), 100253; <https://doi.org/10.1016/j.softx.2019.100253>
 8. Monti, A., Milano, F., Bompard, E., and Guillaud, X. *Converter-Based Dynamics and Control of Modern Power Systems*. Academic Press, 2020.
 9. Šikšnyš, L., Pedersen, T. B., Aftab, M., and Neupane, B. Flexibility Modeling, Management, and Trading in Bottom-up Cellular Energy Systems. In *Proceedings of the 10th ACM Intern. Conf. Future Energy Systems*, June 2019, 170–180; <https://doi.org/10.1145/3307772.3328296>
 10. Steinbrink, C., van der Meer, A. A., Cvetkovic, M., Babazadeh, D., Rohjans, S., Palensky, P., and Lehnhoff, S. Smart grid co-simulation with MOSAIK and HLA: A comparison study. *Computer Science - Research and Development* 33, 1–2, (Feb. 2018), 135–143. Springer International Publishing; <https://doi.org/10.1007/s00450-017-0379-y>

Hartmut Schmeck is a Distinguished Senior Fellow at Karlsruhe Institute of Technology and director at FZI Research Center for Information Technology, Germany.

Antonello Monti is director of the Institute for Automation of Complex Power Systems at RWTH Aachen University and holds a joint appointment with the Fraunhofer Center for Digital Energy, Germany.

Veit Hagenmeyer is director of the Institute for Automation and Applied Informatics at Karlsruhe Institute of Technology and spokesperson of the research program Energy Systems Design of the Helmholtz Association, Germany.

© 2022 ACM 0001-0782/22/4 \$15.00

BY STUART E. MIDDLETON, EMMANUEL LETOUZÉ,
ALI HOSSAINI, AND ADRIANE CHAPMAN

Trust, Regulation, and Human- in-the-Loop AI within the European Region

ARTIFICIAL INTELLIGENCE (AI) systems employ learning algorithms that adapt to their users and environment, with learning either pre-trained or allowed to adapt during deployment. Because AI can optimize its behavior, a unit's factory model behavior can diverge after release, often at the perceived expense of safety, reliability, and human controllability. Since the Industrial Revolution, trust has ultimately resided in regulatory systems set up by governments and standards bodies. Research into human interactions with autonomous machines demonstrates a shift in the locus of trust: we must *trust* non-deterministic systems such as AI to self-regulate, albeit within boundaries. This radical shift is one of the biggest issues facing the deployment of AI in the European region.

Trust has no accepted definition, but Rousseau²⁸ defined it as “a psychological state comprising the

intention to accept vulnerability based upon positive expectations of the intentions or behavior of another.” Trust is an attitude that an agent will behave as expected and can be relied upon to reach its goal. Trust breaks down after an error or a misunderstanding between the agent and the trusting individual. The psychological state of trust in AI is an emergent property of a complex system, usually involving many cycles of design, training, deployment, measurement of performance, regulation, redesign, and retraining.

Trust matters, especially in critical sectors such as healthcare, defense, and security, where duty of care is foremost. Trustworthiness must be planned, rather than an afterthought. We can *trust in AI*, such as when a doctor uses algorithms to screen medical images.²⁰ We can also *trust with AI*, such as when journalists reference a social network algorithm to analyze sources of a news story.³⁷ Growing adoption of AI into institutional systems relies on citizens to trust in these systems and have confidence in the way these systems are designed and regulated.

Regional approaches for managing trust in AI have recently emerged, leading to different regulatory regimes in the U.S., the European region, and China. We review these regulatory divergences. Within the European region, research programs are examining how trust impacts user acceptance of AI. Examples include the UKRI Trustworthy Autonomous Systems Hub,^a the French *Confiance.ai* project,^b and the German AI Breakthrough Hub.^c Europe appears to be developing a “third way,” alongside the U.S. and China.¹⁹

Healthcare contains many examples of AI applications, including online harm risk identification,²⁴ mental health behavior classification,²⁹ and

a <https://www.tas.ac.uk>

b <https://www.confiance.ai>

c <https://breakthrough-hub.ai>



automated blood testing.²² In defense and security, examples include combat management systems⁹ and using machine learning to identify chemical and biological contamination.¹ There is a growing awareness within critical sectors^{15,33} that AI systems need to address a “public trust deficit” by adding reliability to the perception of AI. In the next two sections, we discuss research highlights around the key trends of building safer and more reliable AI systems to engender trust and put humans in the loop with regard to AI systems and teams. We conclude with a discussion about applications, and what we consider the future outlook is for this area.

Recent Changes in the Regulatory Landscape for AI

The E.U. is an early mover in the race to regulate AI, and with the draft E.U. AI Act,^d it has adopted an *assurance-based regulatory environment* using yet-to-be-defined AI assurance standards.

d <https://bit.ly/3FATnNj>

These regulations build upon GDPR data governance and map AI systems into four risk categories. The lowest risk categories self-regulate with transparency obligations. The highest risk categories require first-party or third-party assessments enforced by national authorities. Some applications are banned outright to protect individual rights and vulnerable groups.

The U.K. AI Council AI Roadmap^e outlines a sector-specific *audit-led regulatory environment*, along with principles for governance of AI systems including open data, AI audits, and FAIR (Findable, Accessible, Interoperable, Reusable) principles. An example of sector-specific governance is the U.K. online safety bill,^f which assigns a duty of care to online service providers and mandates formal risk assessments by the U.K. telecom regulator OFCOM.

Outside the European region, the U.S. National Security Commission on

e <https://www.gov.uk/government/publications/ai-roadmap>

f <https://www.gov.uk/government/publications/draft-online-safety-bill>

AI report 2021^g outlined a *market-led regulatory environment*, with government focus areas of robust and reliable AI, human-AI teaming, and a standards-led approach^h to testing, evaluation, and validation. China’s AI development plan²⁷ emphasizes societal responsibility; companies chosen by the Chinese state to be AI champions follow national strategic aims, and state institutions determine the ethical, privacy, and trust frameworks around AI.

The European region, driven by U.K. and E.U. AI regulation, is creating a “third way” alongside the AI regulation adopted by the U.S. and China. This “third way” is characterized by a strong European ethical stance around AI applications, for example limiting the autonomy of military AI systems, in direct contrast to China, where autonomy for AI-directed weapons is actively encouraged as part of its military-civil fusion strategy.¹⁴ It also is characterized by a strong European

g <https://www.nsc.ai.gov/2021-final-report>
h <https://www.nist.gov>

The E.U. is an early mover in the race to regulate AI, and with the draft E.U. AI Act, it has adopted an assurance-based regulatory environment using yet-to-be-defined AI assurance standards.

focus on a citizen's right to data privacy and the limits set on secondary data processing by AI applications, in contrast to China and the U.S., where state-sponsored strategic aims or weak commercial self-regulation around AI applications frequently override data privacy concerns. An example of this "third way" in action is the European city of Vienna becoming the first city in the world to earn the IEEE AI Ethics Certification Mark,³⁰ which sets standards for transparency, accountability, algorithmic bias, and privacy of AI products. How different regional approaches to AI regulation perform in the heat of geo-political AI competition is likely to shape how regional AI research is conducted for many years to come.

Building Safe and Reliable AI to Engender Trust

Assuring safe, reliable AI systems can provide a pathway to trust. However, non-deterministic AI systems require more than just the application of quality assurance protocols designed for conventional software systems in well-regulated regions such as Europe. New methods are emerging for the assurance of the machine learning life cycle from data management to model learning and deployment.²

Exploratory data analysis and *adversarial generative networks* help assure training data comes from a trusted source, is fit for the purpose, and is unbiased. *Built-in test (BIT)* techniques support model deployment, such as watchdog timers or behavioral monitors, as well as "last safe" *model checkpointing* and *explainable AI* methods. Active research focuses on explainable machine learning.⁵ Approaches include *explanation by simplification*, such as local interpretable model-agnostic explanations (LIME) and counterfactual explanations; *feature relevance techniques*, such as Shapley Additive Explanations (SHAP) and analysis of random feature permutations; *contextual and visual explanation* methods such as sensitivity analysis and partial dependence plots; and full life-cycle approaches such as the use of provenance records. Research challenges for assurance of machine learning include detection of problems before critical failures, con-

tinuous assurance of adaptive models, and assessing levels of independence when multiple models are trained on common data.

The manufacturing sector and smart cities deployments increasingly are using digital twins,³⁶ simulations of operating environments, to provide pre-deployment assurance. Digital twins also are used in healthcare,⁸ for example to assure pre-surgical practice, and other critical sectors. A recent U.K.-hosted RUSI-TAS Conference³⁵ discussed how digital twins can provide AI models with a safe space to fail. Other research trends include probing vulnerabilities of AI to accidents or malicious use. This includes examining how malicious actors can exploit AI.¹¹ Attack vectors include adversarial inputs, data poisoning, and model stealing. Possible solutions include safety checklists¹² and analysis of hostile agents that use AI to subvert democracies.³¹

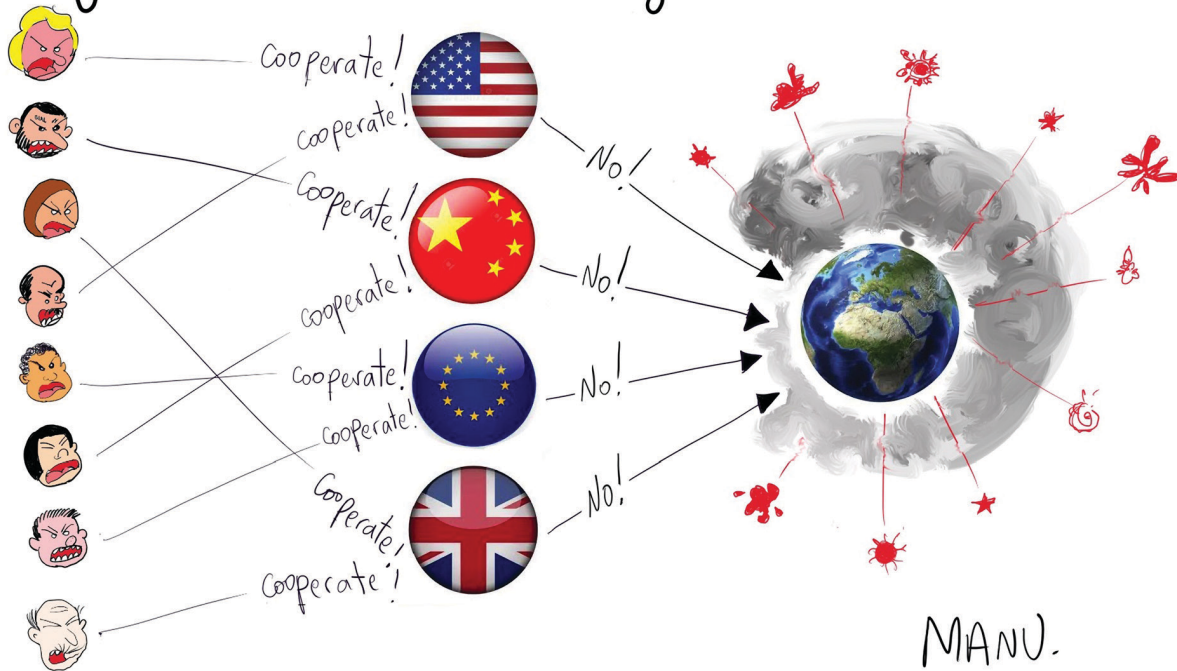
Safe and Reliable AI has received a lot of attention in the European region recently compared to the U.S. and China, and it is no coincidence that every one of the works cited in this section are from authors based in this region. This level of activity is probably motivated by the assurance and audit-based European regulatory stances. The more we understand the vulnerabilities and assurance protocols of AI, the safer and more reliable AI systems will become. Safe, transparent systems that address user concerns will encourage public trust.

Human and Society in the Loop

Human-in-the-Loop (HITL) systems are grounded in the belief that human-machine teams offer superior results, building trust by inserting human oversight into the AI life cycle. One example is when humans mark false positives in email spam filters. HITL enhances trust in AI by optimizing performance, augmenting data, and increasing safety. It enhances trust by providing transparency and accountability: unlike many deep learning systems, humans can explain their decisions in natural language.

However, the AI powering social media, commerce, and other activities may erode trust and even sow discord.⁴ If perceived as top-down oversight from

If the world were a giant neural network



experts, HITL is unlikely to address public trust deficits. Society-in-the-Loop (SITL) seeks broader consensus by extending HITL methods to larger demographics,^{16,25} for instance by crowdsourcing the ethics of autonomous vehicles to hundreds of thousands of people. Another approach is co-design with marginalized stakeholders. The same imperative drives CODES (Council for the Orientation of Development and Ethics) in AI and data-driven projects in developing countries,ⁱ where representatives of local stakeholder groups provide feedback during project life cycles. SITL combined with mass data literacy⁷ may reweave the fabric of human trust in and with AI.

A growing trend is to add humans into deep learning development and training cycles. Human stakeholders *co-design* AI algorithms to encourage responsible research innovation (RRI), embed end-user values, and consider the potential for misuse. During AI training, traditional methods such as *adversarial training* and *active learning* are applied to the deep learning models^{13,21} using humans to label

uncertain or subjective data points during training cycles. *Interactive sense making*¹⁷ and *explainable AI*⁵ also can enhance trust by visualizing AI outputs to reveal training bias, model error, and uncertainty.

Research into HITL is much more evenly spread across the European, U.S., and Chinese regions than work on safe and reliable AI, with about half the work cited in this section from authors based in the European region. Where the European region does differentiate itself is with a stronger focus on HITL to promote ethical AI and responsible innovation, as opposed to the U.S. and China, where there is a tighter focus on using HITL to increase AI performance.

Applications in Critical Sectors

AI offers considerable promise in the following sectors. Each illustrates high-risk, high-reward scenarios where trust is critical to public acceptance.

Defense. General Sir Patrick Sanders, head of U.K. Strategic Command, recently emphasized, “Even the best human operator cannot defend against multiple machines making thousands of maneuvers per second

at hypersonic speeds and orchestrated by AI across domains.”¹⁸ While human-machine teaming dominates much current military thinking, by taking humans *out* of the loop AI transforms the tempo of warfare beyond human capacity. From strategic missile strikes to tactical support for soldiers, AI impacts every military domain and, if an opponent has a high tolerance for error, it offers unstoppable advantages. Unless regulated by treaty, future warriors and their leaders will likely trust AI as a matter of necessity.

Law enforcement and security. Law enforcement is more nuanced. Though used only for warnings, Singapore’s police robots have provoked revulsion in European press,³⁴ and the E.U. AI Act reflects this attitude by classifying law enforcement as high-risk. Some groups have claimed ambiguities in the E.U. AI Act leave the door open for bias, unrestrained surveillance, and other abuses,³² but at minimum it provides a framework for informed progress while asserting the European region’s core values.

Healthcare. Healthcare interventions directly impact lives. Research

i <https://datapopalliance.org>

into diagnostic accuracy shows that AI can improve healthcare outcomes.^{6,10,23,26} However, starting with patients and physicians, trust cascades upward, and as Covid has shown, trust is ultimately political, and thus needs to be nurtured carefully.

Transportation. Self-driving cars may receive the most publicity, but AI also is applied to mass transit, shipping, and trucking. Transportation involves life-or-death decisions, and the introduction of AI is changing the character of liability and assurance. These questions reflect a fundamental question which is being debated today: Who does the public trust to safely operate a vehicle?

Future Outlook


We think future standards for assurance will need to address the non-deterministic nature of autonomous systems. Whether robotic or distributed, AI is effectively an entity, and regulation, management, and marketing will need to account for its capacity to change.

Many projects currently are exploring aspects of bringing humans into the loop for co-design and training of AI systems and human-machine teaming. We think this trend will continue, and if coupled with genuine transparency, especially around admitting AI mistakes and offering understandable explanations for why these mistakes happened, offers a credible pathway to improving the state of public trust in AI systems being deployed into society.

We think that increasingly, *Trust with AI* will shape how citizens trust information, which has the potential to reduce the negative impact of attempts to propagate disinformation. If citizen trust in the fabric of AI used within society is reduced, then *trust in AI* itself will weaken. This is likely to be a major challenge for our generation.

Creating regulatory environments that allow nation-states to gain commercial, military, and social advantages in the global AI race may be the defining geopolitical challenge of this century. Regulation around AI has been developing worldwide, moving from self-assessment guidelines³ to frameworks for national or transna-

tional regulation. We have noted that there are clear differences between the European region and other areas with robust capacity in AI, notably the need for public acceptance. The future will be a highly competitive environment, and regulation must balance the benefits of rapid deployment, the willingness of individuals to trust AI, and the value systems which underlie trust.

Acknowledgments. This work was supported by the Engineering and Physical Sciences Research Council (EP/V00784X/1), Natural Environment Research Council (NE/S015604/1), and Economic and Social Research Council (ES/V011278/1; ES/R003254/1). 

References

- Alan Turing Institute. Data Study Group Final Report: DSTL—Anthrax and nerve agent detector. (2021); <https://doi.org/10.5281/zenodo.4534218>
- Ashmore, R., Calinescu, R., and Paterson, C. Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *ACM Comput. Surv.* 54, 5 (2021), Article 111; <https://doi.org/10.1145/3453444>
- Ayling, J. and Chapman, A. Putting AI ethics to work: Are the tools fit for purpose? *AI Ethics* (2021); <https://doi.org/10.1145/3453444>
- Barrett, P., Hendrix, J., and Sims, G. How tech platforms fuel U.S. political polarization and what government can do about it. The Brookings Institution 27 (Nov. 2021); <https://brookings.org/3sk3Cev>
- Belle, V., and Papantonis, I. Principles and Practice of Explainable Machine Learning. (2020); [arXiv:2009.11698](https://arxiv.org/abs/2009.11698)
- Bhandari, M., Zeffiro, T., and Reddiboina, M. Artificial intelligence and robotic surgery: Current perspective and future directions. *Curr Opin Urol.* 30, 1 (2020), 48–54; [doi:10.1097/MOU.0000000000000692](https://doi.org/10.1097/MOU.0000000000000692)
- Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., and Shoup, N. Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data. Data-Pop Alliance White Paper. Sept. 29, 2015; <https://bit.ly/3qNgBtm>
- Bruynseels, K., Santoni de Sio, F., and van den Hoven, J. Digital twins in health care: Ethical implications of an emerging engineering paradigm. *Front. Genet.* 9, 31 (2018); <https://doi.org/10.3389/fgene.2018.00031>
- AI and data science: Defense science and technology capability, Aug. 1, 2021; <https://www.gov.uk/guidance/ai-and-data-science-defence-science-and-technology-capability>
- Gumbs, A.A., Frigerio, I., Spolverato, G., Croner, R., Illanes, A., Chouillard, E., and Elyan, E. Artificial intelligence surgery: How do we get to autonomous actions in surgery? *Sensors (Basel)* 21, 16 (2021); <https://www.mdpi.com/1424-8220/21/16/5526>
- Hartmann, K., Steup, C. Hacking the AI—The next generation of hijacked systems. In *Proceedings of the 12th Intern. Conf. Cyber Conflict*, 2020, 327–349; [doi:10.23919/CyCon49761.2020.9131724](https://doi.org/10.23919/CyCon49761.2020.9131724)
- Hunt, E.R., and Hauert, S. A checklist for safe robot swarms. *Nature Machine Intelligence* (2020); [doi:10.1038/s42256-020-0213-2](https://doi.org/10.1038/s42256-020-0213-2)
- Kanchinadam T., Westpfahl, K., You, Q., and Fung, G. Rationale-based human-in-the-loop via supervised attention. In *Proceedings of 1st Workshop on Data Science with Human in the Loop* (Aug. 24, 2020); <https://bit.ly/3eYKJJA>
- Kania, E.B. Chinese military innovation in the AI revolution. *The RUSI J* 164, 5–6 (2019), 26–34; [doi:10.1080/03071847.2019.1693803](https://doi.org/10.1080/03071847.2019.1693803)
- Kerasidou, C., Kerasidou, A., Buscher, M., and Wilkinson, S. Before and beyond trust: Reliance in medical AI. *J Medical Ethics* (2021); <http://dx.doi.org/10.1136/medethics-2020-107095>
- Larsson, S. The socio-legal relevance of artificial intelligence. *Droit et société* 103, (2019) 573–593; [doi:10.3917/drs1.103.0573](https://doi.org/10.3917/drs1.103.0573)
- Middleton, S.E., Lavorgna, L., Neumann, G., and Whitehead, D. Information extraction from the long tail: A socio-technical AI approach for criminology investigations into the online illegal plant trade. *WebSci '20 Companion*, 2020.
- Ministry of Defense. Commander of Strategic Command RUSI conference speech, May 26, 2021; <https://www.gov.uk/government/speeches/commander-of-strategic-command-rusi-conference-speech>
- Morton, S. and Booth, M. The EU's "third way" to AI regulation. Pillsbury, Sept. 22, 2021; <https://www.internetandtechnologylaw.com/eu-third-way-ai-regulation/>
- NHS-X. Cancer digital playbook. (2021); <https://www.nhs.uk/keys/key-tools-and-info/digital-playbooks/cancer-digital-playbook/>
- Nie, Y., Williams, A., Dinan, E., Bansal, M., Weston, J., and Kiela, D. Adversarial NLI: A new benchmark for natural language understanding. *ACL* (2020)
- Pinpoint. Early Cancer Detection. (2021); <https://www.pinpointdatascience.com>
- Prabhakar, B., Singh, R.K., and Yadav, K.S. Artificial intelligence (AI) impacting diagnosis of glaucoma and understanding the regulatory aspects of AI-based software as medical device. *Computerized Medical Imaging and Graphics* 87 (2021).
- ProTechThem. ESRC grant ES/V011278/1 (2021); <http://www.protechthem.org>
- Rahwan, I. Society-in-the-loop: programming the algorithmic social contract. *Ethics Inf Technol* 20, (2018), 5–14; [doi:10.1007/s10676-017-9430-8](https://doi.org/10.1007/s10676-017-9430-8)
- Rangarajan, A.K., Ramachandran, H.K. A preliminary analysis of AI-based smartphone application for diagnosis of COVID-19 using chest X-ray image. *Expert Systems with Applications* 183, (2021); [doi:10.1016/j.eswa.2021.115401](https://doi.org/10.1016/j.eswa.2021.115401)
- Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., and Floridi, L. The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & Soc* 36 (2021) 59–77; [doi:10.1007/s00146-020-00992-2](https://doi.org/10.1007/s00146-020-00992-2)
- Rousseau, D.M., Sitkin, S.B., Burt, R.S., and Camerer, C. Not so different after all: A cross-discipline view of trust. *Academy of Management Rev.* 23, (1998), 393–404; [doi:10.5465/AMR.1998.926617](https://doi.org/10.5465/AMR.1998.926617)
- SafeSpacesNLP, UKRI TAS agile project, (2021); <https://www.tas.ac.uk/safespacesnlp>
- Schabus, D. The IEEE CertifAIEd Framework for AI Ethics Applied to the City of Vienna. IEEE Standards Assoc. (2021); <https://bit.ly/3EZWzRp>
- Schia, N.N. and Gjesvik, L. Hacking democracy: managing influence campaigns and disinformation in the digital age. *J. Cyber Policy* 5, 3 (2020), 413–428; [doi:10.1080/23738871.2020.1820060](https://doi.org/10.1080/23738871.2020.1820060)
- Skelton, S.K. NGO Fair Trials calls on E.U. to ban predictive policing systems. *ComputerWeekly* (2021); <https://bit.ly/3mG4uWV>
- Taddeo, M., McCutcheon, T., and Floridi, L. Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nat Mach Intell* 1, (2019), 557–560; [doi:10.1038/s42256-019-0109-1](https://doi.org/10.1038/s42256-019-0109-1)
- The Guardian* via Agence France Presse. 'Dystopian world': Singapore patrol robots stoke fears of surveillance state; <https://bit.ly/3EBQRou>
- RUSI-TAS. Trusting Machines? *RUSI-TAS 2021 Conf.*; <https://www.tas.ac.uk/eventslist/trusting-machines/trust-machines-conference-programme/>
- van der Valk, H., Haße, H., Möller, F., Arbter, M., Henning, J., and Otto, B. A Taxonomy of digital twins. In *Proceedings of AMCIS 2020*; <https://bit.ly/3HE6ZYI>
- European Commission, Horizon 2020 grant agreement 825297 (2021); <https://cordis.europa.eu/project/id/825297>

Stuart E. Middleton is a lecturer in computer science at the University of Southampton, Southampton, UK.

Emmanuel Letouzé is Marie Curie Fellow at Universitat Pompeu Fabra, Barcelona, Spain.

Ali Hossaini is Senior Visiting Research Fellow at Kings College London, UK.

Adriane Chapman is a professor in computer science at the University of Southampton, Southampton, UK.

BY TOMMASO DI NOIA, NAVA TINTAREV,
PANAGIOTA FATOUROU, AND MARKUS SCHEDL

Recommender Systems under European AI Regulations

THE EUROPEAN COMMISSION (EC) has acknowledged the importance artificial intelligence (AI) plays in forming Europe's future, identifying AI as the most strategic technology of the 21st century.^a With a recent proposal on a *Regulation Laying Down Harmonised Rules on Artificial Intelligence*^b (EU Regulatory

Framework for AI), the EC aims at introducing the first comprehensive legal framework on AI, which will identify specific risks for AI, provide a collection of high-risk application domains, propose specific requirements that AI systems should meet when used in such domains, and define obligations for users and providers (U.S. regulatory development relating to AI^c). What clearly emerges from these efforts is the need for an AI that behaves in a responsible way. A clear and globally accepted definition of responsibility for AI systems is still under development, but will likely include notions such as fairness, security and privacy, explainability, safety, and reproducibility. Although safety and reproducibility are fundamental issues in AI research and its industrial application, we will not

cover them here since they are requirements in many areas of technology, therefore not specific to AI.

According to the EC regulation, AI should be used in compliance with the E.U. Charter of Fundamental Rights,^d including the right not to be discriminated against, the respect for private life, and the protection of personal data. The regulation also stresses the “obligations for ex ante testing, risk management and human oversight of AI systems to minimize the risk of erroneous or biased AI-assisted decisions in critical areas such as education and training, employment, important services, law enforcement and the judiciary.” High-risk AI systems should meet specific legal requirements in relation to data management, documentation, human oversight, transparency, robustness, accuracy, and security. According to Article 10, “training, validation and

a See <https://bit.ly/3HTQMP3>

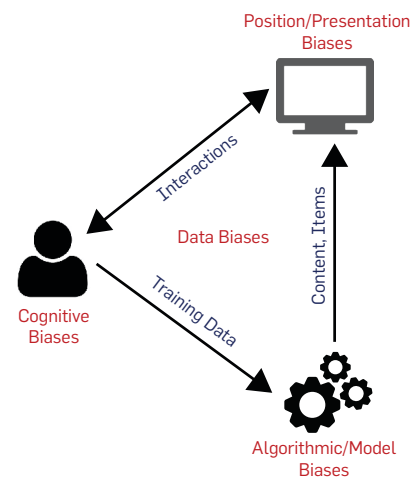
b See <https://bit.ly/34vooEz>

c See <https://bit.ly/3rc2DkO>

d See <https://bit.ly/3r3mrH8>

While much research has been devoted to uncover and mitigate biases in recommender systems, many research gaps still exist.

Figure 1. Different categories of biases and their interplay.



testing data sets shall be subject to appropriate data governance and management practices” which shall concern, in particular, “examination in view of possible biases” and “identification of any possible data gaps or shortcomings, and how those gaps and shortcomings can be addressed.” On the other hand, Article 15 is devoted to accuracy, robustness, and cybersecurity: high-risk AI systems must achieve all three throughout their entire life cycle to a satisfactory degree based on state-of-the-art security and privacy-preserving measures. The regulation makes it also clear that “AI systems should be sufficiently transparent, explainable and well-documented” (Article 13).

In the following, we attempt to provide the European scene for fairness, security and privacy, and explainability under the lens of recommender systems (RSs). Given their user-centric nature, RSs are fully touched by the principles and rules stated in the aforementioned EC documents, and therefore represent an interesting workbench to study their application. Issues related to fairness, security and privacy, and explainability may affect a RS at training and runtime.

Fair Recommender Systems

Despite many EC-proposed provisions regarding AI fairness, in reality, RSs have been shown to provide different recommendation quality to different users, depending on various characteristics, such as gender, age, ethnicity, or personality.^{3,7,8,10-12} Such behavior con-

flicts with the aforementioned goals and is likely to yield unfairness.

Definitions. It is common to distinguish between *individual fairness* and *group fairness*. The former means that similar users are treated in a similar fashion (for example, users with similar skills receive job recommendations within the same pay grade). The latter means that different groups of users defined by some sensitive or protected attribute (for example, gender or ethnicity) are treated in the same way. Accordingly, unfairness is defined as “systematically and unfairly discriminat[ing] against certain individuals or groups of individuals in favor of others.”⁵

Categories of biases. Unfairness is commonly caused by societal or statistical biases, the former referring to the divergence between how the world should be and how it actually is, the latter to the discrepancy between how the world is and how it is encoded in the system. Such biases can occur at different levels in the recommendation pipeline (see Figure 1). They can be present already in the *data* the algorithms are trained on (for example, an unbalanced dataset with respect to representation of different genders), they can be amplified by the *algorithms* or created *models* (for example, reinforcing stereotypes), or they can originate from users, *cognitive biases* (for example, serial position, anchoring, and decoy effects).⁸

Bias mitigation strategies.

To alleviate existing biases, several techniques can be adopted. Focusing on data and algorithm/model bias, the most common approaches are *data rebalancing* (for example, upsampling the minority group of users in the dataset), *regularization* (for example, including a bias correction term in the loss function of the machine/deep learning algorithm), and *adversarial learning* (for example, training a classifier that tries to predict the sensitive attribute from the user-item interaction data and modify the data or recommendation algorithm to minimize the classifier’s accuracy).

While much research has been devoted to uncover and mitigate biases in RSs, both within and outside the E.U., many research gaps, the most pressing ones including:

- Several metrics of fairness have been proposed. However, a compre-

hensive (formal and comparative) study of their strengths and limitations is still missing.^e Even whether they reflect what humans perceive as fair or unfair—possibly depending on their cultural background, values, and beliefs—has not yet been investigated deeply.

► Likewise, a thorough understanding of capabilities and limitations of existing techniques for mitigating bias through their systematic evaluation is missing.

► From an algorithmic perspective, novel methodologies to debias state-of-the-art RS algorithms, which are predominantly based on deep learning, are needed.

► An investigation of potential economic and social consequences of biases resulting from the use of RSs adopted in high-risk areas (for example, in recruitment) is needed.^{3,4}

► Fairness is typically addressed

from a system's end user's perspective, but we need to consider multiple RS stakeholders, including content producers, content consumers, and platform providers.

► From a legal perspective, we need provisions with respect to data quality, concrete specifications to whom the obligations to not violate EU non-discrimination law applies, and effective mechanisms for auditing RSs for legal compliance. This requires an interdisciplinary perspective, involving collaboration between researchers with technical expertise and law experts.

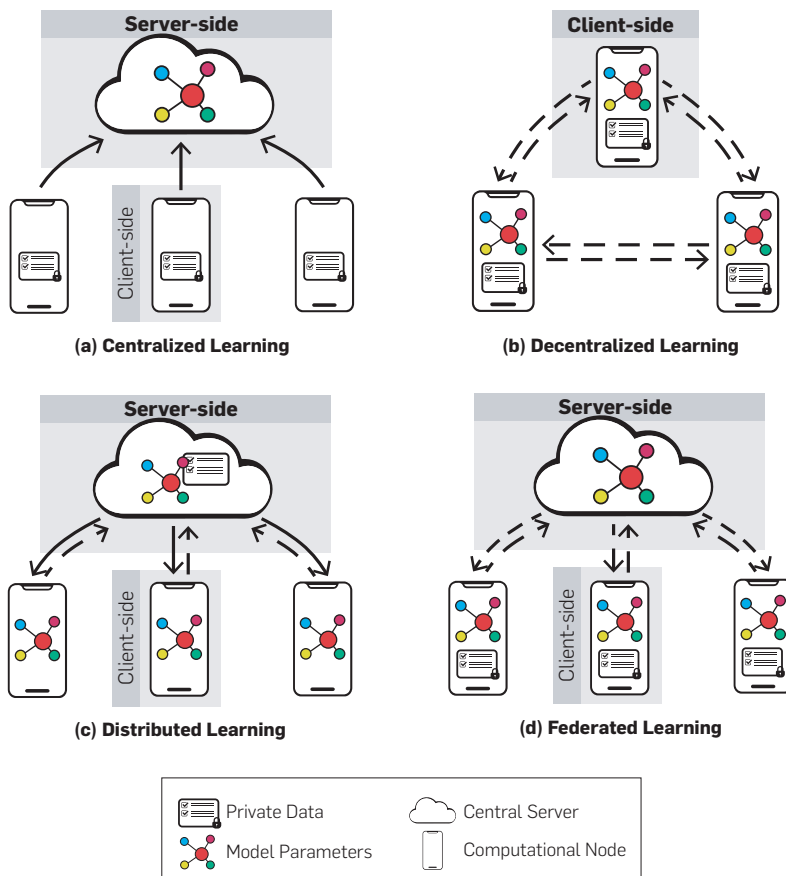
Security and Privacy for Recommender Systems


Privacy in AI is a dominant concern in the EU. To comply with GDPR, AI strategies must be applied considering the new privacy challenges that may limit the uptake of these applications. Privacy-related risks are even more evident when we think about the applications of RSs where user models are

Together with attack strategies, defense mechanisms against adversarial attacks have been developed in recent years.


e See <https://bit.ly/3zPOFZK>

Figure 2. Information flow over the network in four ML architectures. Solid lines represent training data flow, dashed lines represent model parameters flow.





Human oversight is not feasible if explanations are not understandable by people.



built around personal data.

Issues. Data fragmentation and isolation while complying with the GDPR is a major challenge for RS researchers. Actually, preserving users' privacy is not as easy as limiting data collection since a privacy threat may happen at any stage of the data cycle. The model itself stores precious information able to predict future user preferences and behaviors. The main target of privacy attacks in a RS is the confidentiality of the users' sensitive data. Privacy-preserving ML aims to equip ML with defense measures for protecting user privacy and data security. It should be distinguished from secure ML, which attempts instead to preserve integrity and availability of a ML system from intentional (adversarial or poisoning) attacks.

Federated learning. From a privacy perspective, federated learning (FL)⁹ completely addresses the principles of focused collection, data minimization, data ownership, and data locality, thus greatly reducing the privacy risks of centralized learning (see Figure 2). While handling users' privacy concerns, FL faces challenges such as communication costs, unbalanced data distribution and device reliability, and security issues (for example, model poisoning, indirect information leakage, and Byzantine adversaries). In Yang et al.,¹⁵ the concept of FL is extended to a more comprehensive idea of privacy-preserving decentralized collaborative ML techniques, both for horizontal federations (where different datasets share the same feature space but are different in training samples) and vertical federations (where different datasets share the training samples but differ in feature space). Thanks to tunable federation approaches in recommendation scenarios, users can be more aware of and decide which data they share.¹

Unfortunately, while FL can offer significant practical privacy improvements over centralized approaches, there is still no formal guarantee of privacy. This is where other techniques, such as differential privacy, secure multiparty computation, and homomorphic encryption, come to the stage to enforce the privacy protection mechanisms also in recommendation scenarios.

Adversarial attacks and defense.

Notwithstanding the great success of machine/deep learning models, recent studies have shown they are not immune to security threats from adversarial use of AI, and the same holds for RSs.² An adversary can attack a ML model at two main stages of the learning pipeline, during training or production. These two categories of attacks are respectively known as training-time attack (a.k.a. causative or poisoning attack) and inference-time attack (a.k.a. exploratory or evasion attack).

► **Poisoning attack.** Data poisoning attacks are realized by injecting false data points into the training data with the goal to corrupt/degrade the model (for example, the classifier).

► **Evasion attack.** Instead of interfering with training data, evasion attacks adjust malicious samples during the inference phase. These attacks are also named *decision-time attacks* referring to their attempt to evade the decision made by the learned model at test time.

Adversarial examples created for image classification tasks are empowered based on continuous real-valued representation of pixels, but in RSs the raw values are user/item identifiers and ratings that are discrete. Hence, adversarial perturbations are added to: the user profile directly (that is, user rating profile), user and item model parameters in a latent factor model; and embeddings representing side information of user and items, respectively.

Together with attack strategies, defense mechanisms against adversarial attacks have been developed in recent years. They can be classified into detection methods and methods seeking to increase the robustness of the learning model. At the heart of the robust optimization method is the assumption that every sample in the training data can be a source for adversarial behavior. It applies a zero-sum game between the prediction and attack adversaries. The ultimate goal in robust optimization is that the prediction model will perform equally well with adversarial and clean inputs.

Explainable Recommender Systems

Although RSs operate as artificial advice givers, people using the system may not understand how the

conclusion was reached and when it is appropriate to adopt the advice, or in contrast, when to critique it. The *EU Regulatory Framework for AI* consequently indicates that explanations need to supply human oversight of high-risk systems.

There are three main scientific challenges to overcome before compliance is possible, namely: how to ensure explanations are human-understandable and support appropriate trust; how to build explainable AI (explanation confidence and model complexity); and how to validate the goodness of explanations.

Understandability. Human oversight is not feasible if explanations are not understandable by people, and “*Interpretability*” has been qualified as the degree to which a human can understand the cause of a decision.¹³ Understanding is rarely an end-goal in itself, and it is often more useful to measure the effectiveness of explanations in terms of a specific notion of usefulness or explanatory goals such as improved decision support or (appropriate) user trust¹⁴—minimizing both over- and underreliance on system advice. Furthermore, both the characteristics of the people (for example, expertise, cognitive capacity) and the situation (such as which other people are affected, or which variables are influential) place different requirements on which explanations are useful, also to different presentational choices (for example, with regard to modality, degree of interactivity, level of detail). Simply put: One size does *not* fit all.

Building explainable AI (XAI). A large number of methods for XAI have been developed, for a breadth of models and types of data. However, many of them do not (by design) support users in fully understanding the capacities and limitations in a way that would support appropriate trust. We identify two particularly limiting barriers: limited model confidence and high model complexity.

► **Confidence.** For sufficient human oversight, RSs must be aware of their knowledge limits not only on the prediction (global and instance) level but also on the explanation level. Consequently, RSs must provide confidence information for each prediction and explanation; and they must clarify how

this information has been obtained or computed.

► **Complexity.** While it is commonly (but erroneously) believed there is a trade-off between accuracy and interpretability, this is not strictly true. In many cases, several models can offer comparable accuracy performance, but some are more human-understandable. Complexity can be mitigated by selecting the simpler model, and by developing interactive interfaces such as those we have developed in our work, which: adapt the generated explanations to different factors and allow people using the system to see how the factors influence the explanations (transparency), as well as modify the contribution of the factors (control).⁶

Evaluation of explanations. User studies are indispensable for evaluating human performance and understanding. However, to date they are relatively rare in the literature, likely due to their cost. Explanations have also been subjected to automated evaluations, modeled as penalties and constraints in optimization problems, sparsity of the model, monotonicity with respect to a variable, or decomposability into sub-models, and so forth. However, so far there have been no standardized metrics developed. As for ML (Precision, Recall, F-measure, AUC), perhaps this is also because there is no single ideal objective (for example, accuracy versus succinctness). Nevertheless, we hope in the coming years to see benchmarking of such metrics as we see in challenges such as Kaggle, CLEF, and SemEval.

Conclusion

The wide adoption of AI algorithms and systems calls for the definition and realization of a responsible approach to AI. In this respect, by following the documents and legal frameworks proposed over the last years by the EC, some technological issues and trends emerge. We require AI systems to be fair, secure, and privacy-preserving, and interpretable. In this article, we outlined the steps that have already been taken in this direction, as well as indicating what we see as the challenges ahead before we can fulfill the spirit of the European approach to AI. We hope this will serve as a useful roadmap for practitioners and researchers alike.

Acknowledgments. This work

has been supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement numbers 860621 and 101031688, by the Linz Institute of Technology, project “Mitigating Gender Bias in Job Recommender Systems: A Machine Learning-Law Synergy” (TIMELY), and by the projects “Casa delle Tecnologie Emergenti della Città di Matera,” H2020-ICT-2018-20 “Passepartout,” and “Safe and Secure Apulia.” T. Di Noia wishes to thank V.W. Anelli, Y. Deldjoo, A. Ferrara, and F. Merra. 

References

1. Anelli, V.W. et al. FedeRank: User controlled feedback with federated recommender systems. *ECIR 1* (2021), 32–47.
2. Deldjoo, Y., Di Noia, T., and Merra, F.A. A Survey on adversarial recommender systems: From attack/defense strategies to generative adversarial networks. *ACM Comput. Surv.* 54, 2, Article 35 (Mar. 2022).
3. Fatourou, P., Hankin, C., and Knowles, B. Gender Bias in Automated Decision Making Systems. Policy paper, endorsed by the ACM Europe Technology Policy Committee (2021); <https://bit.ly/3r7EHik>
4. Fatourou, P., Papageorgiou, Y., and Petoussi, V. Women are needed in STEM: European policies and incentives. *Commun. ACM* 62, 4 (Apr. 2019), 52.
5. Friedman, B. and Nissenbaum, H. Bias in computer systems. *ACM Trans. Inf. Syst.* 14, 3 (July 1996), 330–347.
6. Jin, Y. et al. Effects of personal characteristics in control-oriented user interfaces for music recommender systems. *User Modeling and User-Adapted Interaction* 30, 2 (2020), 199–249.
7. Lambrecht, A. and Tucker, C.E. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Manag. Sci.* 65, 7 (2019), 2966–2981.
8. Lex, E. Psychology-informed recommender systems. *Found. Trends Inf. Retr.* 15, 2 (2021), 134–242.
9. McMahan, B. et al. Communication-efficient learning of deep networks from decentralized data. *AISTATS 2017*, 1273–1282.
10. Mansoury, M. et al. Feedback loop and bias amplification in recommender systems. In *Proceedings of the 29th ACM Intern. Conf. Information & Knowledge Management* (2020), 2145–2148.
11. Melchiorre, A.B., Zangerte, E. and Schedl, M. Personality bias of music recommendation algorithms. In *Proceedings of the 14th ACM Conf. Recommender Systems Virtual* (Sept. 2020).
12. Melchiorre, A.B. et al. Investigating gender fairness of recommendation algorithms in the music domain. *Inf. Process. Manag.* 58 (2021).
13. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38.
14. Tintarev, N. and Masthoff, J. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*. Springer (2015), 353–382.
15. Yang, Q. et al. Federated machine learning: Concept and applications. *ACM TIST 10.2* (2019), 12:1–12:19. doi: 10.1145/3298981.

Tommaso Di Noia is a professor of computer science at Politecnico di Bari, Italy.

Nava Tintarev is a professor of Explainable AI at Maastricht University, and a visiting professor at TU Delft, The Netherlands.

Panagiota Fatourou is a professor of computer science at the University of Crete, Greece.

Markus Schedl is a professor at Johannes Kepler University in Linz, Austria, where he also heads the Human-Centered Artificial Intelligence group at the Linz Institute of Technology AI Lab.

BY MOR PELEG, YUVAL SHAHAR, AND SILVANA QUAGLINI

MobiGuide: Guiding Clinicians and Chronic Patients Anytime, Anywhere

THE TREND FOR an aging population, which is typical for Europe and for other high-income regions, brings with it a sharp increase in the number of chronic patients and a shortage of clinicians and hospital beds. Evidence-based clinical decision-support systems are one of the promising solutions for this problem.¹⁵

In the 1990s, different research groups started to develop computer-interpretable clinical guidelines (CIGs)⁷ as a form of evidence-based decision-support systems (DSS). Narrative evidence-based clinical guidelines, focused on a single disease, and containing

recommendations for the disease diagnosis and management, were manually represented in CIG formalisms, such as Asbru,¹¹ GLIF,¹ or PROforma.³ The CIGs formed a network of clinical decisions and actions and served as a knowledge base. The DSS would enact the CIG over a patient's data, entered manually or taken directly from the *electronic health record* (EHR). The patient-specific recommendations would be delivered to the clinician during patient encounters.

The European MobiGuide project⁸ asked the following question: What do chronic patients want? Patients want to live normally. They want to focus on their lives and not on their disease. They do not want to go into the hospitals for long monitoring sessions but remain in their natural environments and lead their normal lives safely. Hence, we were motivated to develop a generic architecture that could support chronic patients and their clinicians.

The MobiGuide architecture is shown in Figure 1. Patients received a body-area network of mobile sensors (for example, a glucometer, a blood pressure sensor, an ECG belt) communicating via Bluetooth with a smartphone. The sensors' biosignals were semantically integrated with hospital EHR data, patient reported outcomes and symptoms, and patient-specific DSS recommendations into a secure *Patient Health Record* (PHR) that followed the HL7 Virtual Medical Record standard. The DSS in MobiGuide was distributed between a full-fledged backend DSS and a local *mobile DSS* (mDSS). The Backend DSS had access to the full PHR and to the full CIG knowledge-base, represented in the Asbru¹¹ CIG language, and enacted through the PICARD engine.¹³ Based on the patient's context, the Backend DSS *projects*, when necessary, components of the CIG knowledge to the mDSS.



START

PROFILE

46

73

37.2

SC 46%

60%

SECURITY

LINK-SPOT

MR-SCAN

BODY TEMP

7:45

7:35

Dec

Nov

Oct

Sep

The MobiGuide architecture includes multiple novelties. First, the distributed decision-support architecture *projects* patient-specific components of the evidence-based CIGs into the *local* mDSS, which can thus operate independently for weeks until it detects predefined [temporal] patterns in the data (embedded within the projected CIG component). Such patterns signal that the patient’s context had significantly changed (for example, a temporal pattern indicating lack of blood-glucose-level control) and allow the mDSS to *call back* the Backend DSS to take control and project a new CIG component.¹⁴ This architecture is quite different from a completely central, a completely distributed, or a traditional client-server architecture.

Second, semantic data integration⁶ is based on HL7 standards, extended to allow interoperability not only of EHR, sensor, or patient-reported data, but also of DSS recommendations that were delivered to patients and clinicians.

While traditional DSSs² deliver advice to clinicians, the third novelty is a focus on patients as end-users. To support patient centrality, the formalized CIGs were also customized by adding *CIG-Customized-Contexts (CCCs)*.⁸ Each CCC (for example, “normal schedule,” “good-glycemic-control”)

defines how the CIG changes for *any* patient that enters this context. Contexts included also social circumstances (such as “living alone”) and preferences regarding options in the guideline whose selection depends on patients’ preferences (for example, a cost-utility trade-off).¹⁰

The fourth novelty was the continuous intelligent data analysis that detected the multivariate temporal patterns in the data, using context-sensitive clinical knowledge and fully exploiting the context-sensitive temporal properties of each clinical concept (for example, their persistence over time within each context), unlike the use of standard laboratory-test cut-off values.

The final innovation was the *generic* architecture, unlike domain-specific architectures. This architecture was validated with CIGs for different diseases and with different sensors: atrial fibrillation patients (with ECG and blood pressure sensors) and gestational diabetes patients (with glucometers and blood pressure sensors).

Scenarios of Using the System

To envision the MobiGuide system from the patient’s perspective, we describe two typical scenarios, reflecting the two patient populations who used the system for up to nine months.

Picture Maria, a 65-year-old atrial fibrillation patient from Italy. Maria told her cardiologist that sometimes, when she is walking her dog she feels anxious because she does not know if she is having an atrial fibrillation event and is not sure if she should take her emergency pill, which she carries with her. Her doctor enrolls her to the MobiGuide system. She receives a smartphone with the MobiGuide app and a mobile ECGs sensor with a belt (as shown in Figure 2) she wears beneath her shirt. When Maria reports her symptoms, she is instructed by the mDSS system to activate the ECG sensor.

In the MobiGuide project, ECG data of patients such as Maria was collected by the sensor and abstracted into one-minute sessions. To increase specificity and prevent false alarms, the DSS’ atrial fibrillation detection algorithm monitored for patterns of two or more sessions with atrial fibrillation events within a 10-minute period. Detected sessions were stored in the PHR. When such sessions were detected for eligible patients (depending on their medication therapy and clinical and social parameters), the DSS also checked that no recommendation for the emergency pill has been delivered in the past four hours. Under these terms, the system recommended to the patient to

Figure 1. Architecture of the MobiGuide DSS.

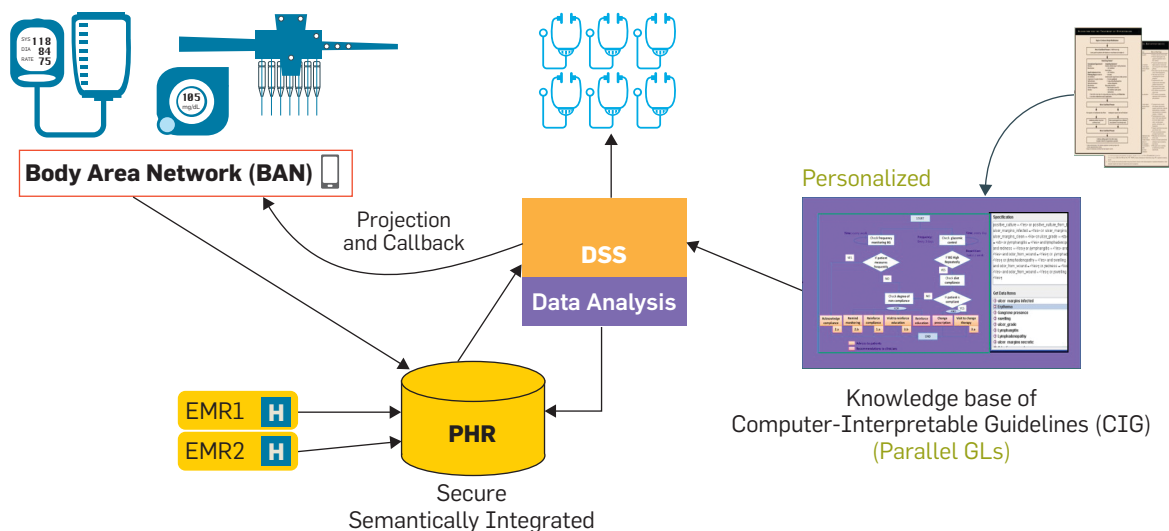


Figure 2. The MobiGuide app; ECG sensor with chest belt, and an atrial fibrillation event detected by the ECG sensor.



rest, take the emergency pill, and then measure the ECG for another 30-minute session. Clinicians could see the data saved into the PHR, and could receive recommendations from the system, supporting a potential cardioversion procedure (such as an ablation procedure).

The second scenario concerns Montse, a young pregnant woman from Spain with gestational diabetes. Montse sets the system in the appropriate context, which currently happens to be her “holiday” context. In this context, she wakes up a bit later, hence her recommendations for measuring blood glucose before breakfast and after her meals are delivered accordingly. At first, she needs to measure her blood glucose four times a day, every day, as her [composite] context is “holiday” and “bad blood-glucose control.” But since her reporting is highly compliant, and since her blood glucose is under control, the system changes the [composite] context to “holiday” and “good blood-glucose control” and advises her to measure her blood-glucose four times a day, only *twice a week*. Hence, the system monitors Montse’s compliance and metabolic control, and accordingly applies evidence-based plans regarding diet, exercise, and measurement schedule, and sends her personalized, context-sensitive reminders.

However, after several weeks, Montse’s blood glucose is too high. The MobiGuide system (following a Call Back to the central server triggered by the blood-glucose values’ pattern, and a projection of a new therapy component) goes back

to recommending daily measurements. In parallel, the system also advises Montse’s managing physician to start insulin. Thus, it asks Montse to see her doctor before the next scheduled visit. The doctor, who had received the MobiGuide system’s recommendation, agrees with it, and starts insulin earlier than she would have done if the system had not alerted her. The MobiGuide system also sends Montse feedback and education, and when needed, advises regarding the addition of carbohydrates if the pre-prandial blood sugar is repeatedly too low.

System Evaluation

MobiGuide was evaluated in a long clinical pilot with 10 atrial fibrillation patients in Italy and 20 gestational diabetes patients in Spain, who were using the system for three to nine months.³

The main benefit to patients was that they stayed at home yet felt better cared for and safer. Their compliance to measurements was high and for the gestational diabetes patients, in which data from a historical cohort were available, a higher compliance of the MobiGuide cohort was observed relative to the historical cohort, reaching 75% for atrial fibrillation patients and 99% for gestational diabetes patients. Relative to the historical cohort, blood pressure was significantly lower in the MobiGuide cohort and there was a trend for fewer C-sections in the MobiGuide cohort, due to the better glycemic control and the lower birth weight of the babies. Moreover, most MobiGuide patients reported an improvement in their quality of life in the EuroQoL questionnaire.

Clinicians used the system even outside patient visits, demonstrating its value to them. Atrial fibrillation clinicians changed a long-time diagnosis for two patients, after realizing from the ECG data stored in the PHR that these patients had in fact other arrhythmias that had been wrongly diagnosed. Gestational diabetes clinicians started insulin earlier for two patients, as recommended by the system, thus

This architecture is quite different from a completely central, a completely distributed, or a traditional client-server architecture.

In the MobiGuide project, ECG data of patients was collected by the sensor and abstracted into one-minute sessions.

enhancing the patients and the embryos' safety. From the clinicians' questionnaire we learned they found the system useful to identify priorities, increase productivity, and they valued the fact that the patient data measured between visits was available to them, which made the patient visits more effective due to availability of data and increased patients' safety.

Discussion

Our experience with the MobiGuide system demonstrated that, given a good mobile network infrastructure, evidence-based guidelines that can be fully disambiguated and formalized, and a motivated clinical and technological supporting team, it is possible to operationalize successfully a large-scale complicated system such as MobiGuide, which features distributed decision support, based on computer-interpretable clinical guidelines, personalized to patient preferences and contexts, and generalizable to different clinical domains. Its successful impact has been evaluated with patients of two types, in two countries, who have used the system for up to nine months.

Managing patients at home implies multiple economic benefits. First, better compliance of patients results in better clinical outcomes. Second, the system can follow up patients remotely, and can thus save unnecessary visits when they are doing well and do not require clinic visits or hospitalization for close monitoring. Moreover, patients who require a faster intervention are detected quickly, enhancing the level of care. (for example, in some cases, the DSS system referred the gestational diabetes patient to the clinic for earlier insulin intervention when conservative measures were deemed to be insufficient).


We believe the MobiGuide architecture is highly scalable; it is generic with respect to *medical knowledge engineering* and with respect to *context-sensitive application*, and thus depends in no way on the particular CIG to be formalized,



as clearly proven in the current evaluation (and in other projects by the team members, in other clinical domains.^{12,13} Its data-security infrastructure has been developed with a large population in mind.

Using innovative systems such as MobiGuide for routine management at home of chronic patients raises multiple intriguing legal issues. First, who is responsible if something goes wrong? Surprisingly, at least at the time the MobiGuide project was developed and tested, the legal situation in the E.U. was that the hardware (smartphone) manufacturer might well be considered as the party at fault if the phone is used to transmit a potentially harmful message. Such a situation might hamper the development and use of systems that can improve patient care and reduce its costs. There are initiatives to move the burden to the software developer, as seems more reasonable. Furthermore, on a more semantic (rather than syntactic) level, specific proposals were made in the past to assess the blame according to the chain of knowledge management and application, from the medical expert, through the knowledge engineer, software developers, sensor developers, and clinicians, and put the blame squarely where it really belongs (for example, a faulty representation of the clinical guideline; or an incomplete application



Acknowledgment. This study was partially funded by the European Commission 7th Framework Program, grant #287811. 

References

1. Boxwala, A.A. et al. GLIF3: A representation format for sharable computer-interpretable clinical practice guidelines. *J. Biomedical Informatics* 37, 3 (2004), 147–161.
2. Doyal, L. Informed consent: Moral necessity or illusion? *Qual Health Care* 10 (Supplement I (2001), i29–i33.
3. Fox, J., Johns, N., and Rahmzadeh, A. Disseminating medical knowledge: The PROforma approach. *Artificial Intelligence in Medicine* 14, 1–2 (1998), 157–182.
4. Kogan, A. Toward a goal-oriented methodology for clinical-guideline-based treatment recommendations for patients with multimorbidity: GoCom and its preliminary evaluation. *J. Biomedical Informatics* (Dec. 2020), 112:103587. doi: 10.1016/j.jbi.2020.103587. Epub 2020 Oct 6.
5. Lisowska, A., Wilk, S., and Peleg, M. Catching patient's attention at the right time to help them undergo behavioural change: Stress classification experiment from blood volume pulse. *Artificial Intelligence in Medicine* (Porto, Portugal, June 16–19, 2021), LNAI 12721.
6. Marcos, C. Solving the interoperability challenge of a distributed complex patient guidance system: A data integrator based on HL7's Virtual Medical Record standard. *J. American Medical Informatics Association* 22, 3 (2015), 587–599.
7. Peleg, M. Computer-interpretable clinical guidelines: A methodological review. *J. Biomedical Informatics* 46, 4 (2013), 744–763.
8. Peleg, M. et al. MobiGuide: a personalized and patient-centric decision-support system and its evaluation in the atrial fibrillation and gestational diabetes domains. *User Modeling and User Adapted Interaction* 27, 2 (2017), 159–213.
9. Peleg, M. et al. Assessment of a personalized and distributed patient guidance system. *Intern. J. Medical Informatics* 101 (2017), 108–130.
10. Quaglini, S. et al. Supporting shared decision making within the MobiGuide Project. In *Proceedings of the Amer. Medical Informatics Assoc. Annual Fall Symp.* (Washington, D.C., 2013), 1175–1184.
11. Shahar, Y., Miksch, S. and Johnson, P.D. The Asgaard project: A task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine* 14, 1–2 (1998), 29–51.
12. Shalom, E. et al. A multiple-scenario assessment of the effect of a continuous-care, guideline-based decision support system on clinicians' compliance to clinical guidelines. *The Intern. J. Medical Informatics* 84, 4 (2015), 248–262.
13. Shalom, E., Shahar, Y., and Lunenfeld, E. An architecture for continuous, user-driven, and data-driven application of clinical guidelines: Addressing the realistic aspects of clinical decision support. *J. Biomedical Informatics* 59 (2016), 130–148.
14. Shalom, E. et al. Distributed application of guideline-based decision support through mobile devices: Implementation and evaluation. *arXiv Preprint*. February 2021. No. 2102.11314. <https://arxiv.org/abs/2102.11314>.
15. Sittig, D.F. et al. Grand challenges in clinical decision support. *J. Biomedical Informatics* 41, 2 (2008), 387–392.

Mor Peleg is a professor in the Department of Information Systems at the University of Haifa in Haifa, Israel.

Yuval Shahar is a professor in the Department of Software and Information Systems Engineering at Ben-Gurion University in Beer-Sheba, Israel.

Silvana Quaglini is a professor in the Department of Electrical, Computer, and Biomedical Engineering at the University of Pavia in Pavia, Italy.

© 2022 ACM 0001-0782/22/4 \$15.00

of a correct representation).

There are also other legal obstacles preventing unhampered use of systems such as MobiGuide throughout Europe (or the world), such as the problem of transporting patient data across borders: in some countries, current national laws often forbid that option, making universal accessibility of personalized care to patients traveling across national borders very tricky.

In addition to legal obstacles, healthcare institutions could raise organizational barriers. As a matter of fact, although systems such as MobiGuide potentially lead to multiple health and economic benefits, their routine implementation does require significant organizational and workflow changes. For example, *care providers* need (re) education, such as how to best exploit evidence-based decision-support systems and not be over-owed by them; also *patients* need (re) education, since patients need to get used to the implications of empowerment, on one hand, but accountability, on the other hand; eventually, human resources need to be reallocated to cope with remote patient monitoring.

Research shows that at least one-third of the patients want empowerment and an informed process of care,² and perhaps the use of such a system (and especially its personalization options) should start with

that patient segment.

To better address patients' needs and maximize the probability of success also beyond the pilot studies, systems like MobiGuide should probably be extended with additional functionalities. One of them is addressing the overall patient's well-being, considering mental well-being, nutrition, exercise, and support for adverse drug effects and multimorbidity. In this sense, the E.U. project Cancer Patients Better Life Experience (CAPABLE; see <https://capable-project.eu>), although based on different technologies, can be seen as an extension of MobiGuide. It will address the well-being of patients by offering also non-medication evidence-based therapies from the mindfulness, positive psychology, and physical exercise domains, grounded in behavior change theories, and fitting interventions to the patient's clinical goals and ability.^{4,5}

Conclusion

MobiGuide is an extraordinary example of an artificial intelligence system that monitors and manages patients remotely, as a solution to the efficient outpatient monitoring in Europe and could provide a solution to the expected shortage of doctors in Europe and around the world. It also inspired further projects that hopefully will overcome some of the highlighted limitations.

BY WIL VAN DER AALST

European Leadership in Process Management

IN THE MID-1990S, many vendors, such as IBM, Staffware, Filenet, Lotus, and Xerox, provided workflow management (WFM) software. In fact, WFM systems were expected to become an integral part of every information system. Despite these high expectations, only a few organizations successfully used this technology. After the limited success of WFM systems, the scope was broadened beyond automation, leading to a wave of business process management (BPM) systems.^{1,7} Many organizations documented their processes using notations such as the business process model notation (BPMN), but few successfully used BPM technologies to create information systems driven by process models.⁸

One of the main reasons was because traditional process management approaches underestimated the complexity and variability of real-world processes and did not explicitly use the data available in existing enterprise resource planning (ERP) and customer relationship management (CRM) systems. Although the first *process-mining algorithms*⁸ were developed around the turn of the century, large-scale adoption of process mining is rather recent. Most considered the WFM/BPM field to be “dead” because model-driven approaches did not live up to their expectations. However, the uptake in process mining has radically changed this notion over the past five years.

According to Gartner, there are now more than 40 process-mining vendors.⁴ Last year the market grew 70% and is expected to grow 40%–50% each year.⁴ Celonis is the leading process-mining vendor and Germany’s first decacorn—that is, a startup valued at more than 10 billion USD. Most process-mining companies are based in Europe—for example, Celonis, LanaLabs, Signavio, MPM, and PAFnow from Germany; ProcessGold, Mavim, and Fluxicon from the Netherlands; Livejourney from France; MyInvenio and Integris from Italy; QPR from Finland; and Minit from Slovakia. Europe’s leadership in this market may be explained by the observation that most process-mining research has been conducted there. Given the focus of this special section of *Communications*, in this article we describe process mining’s European roots as well as developments in the field. Using Celonis as an example, we show how process mining differs from traditional process management and automation.

First, we explain the pitfalls of traditional, purely model-based approaches. Then, we introduce the field of process mining and elaborate on more recent developments, where process mining fuels new forms of automation to address performance and compliance problems. Finally, we focus on Celonis as a successful example of the transfor-



mation from academic research into one of Europe's most successful IT companies.

Model-Driven Process Management

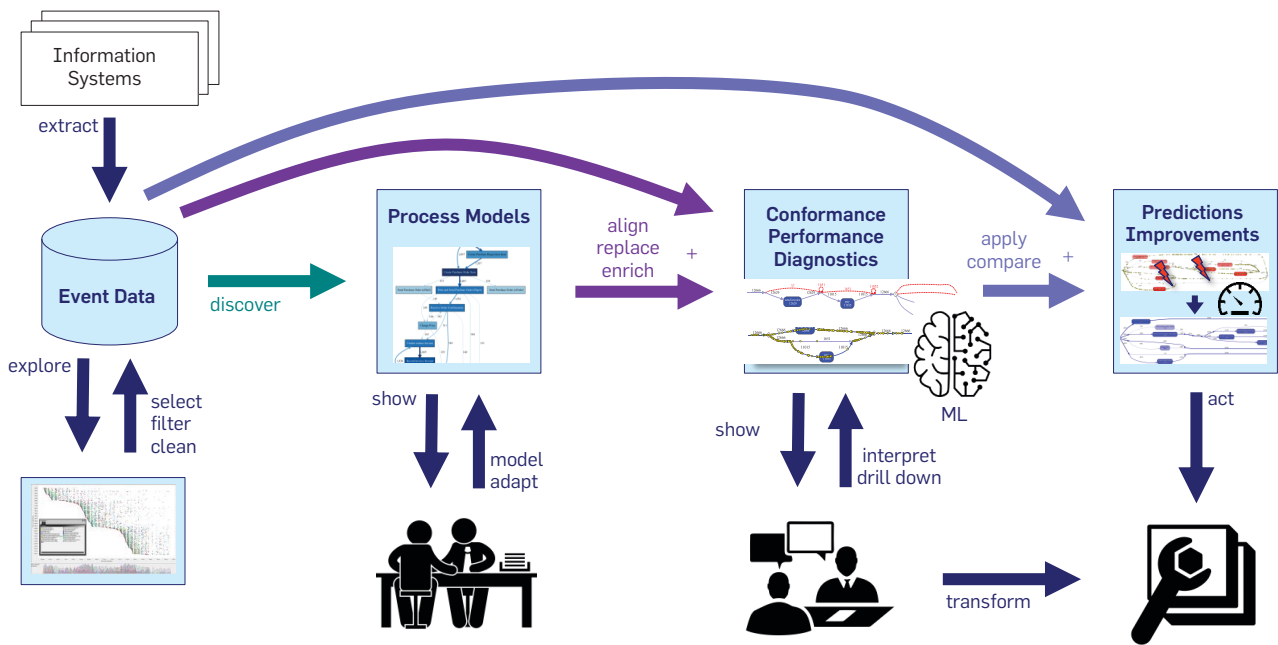
The first prototype *office information systems* were developed in the 1970s. Michael Zisman developed the SCOOP system in the context of his Ph.D. project,¹⁰ and Skip Ellis and colleagues at Xerox developed the OfficeTalk system.² This was when people dreamed of the “office of the future”—that is, completely paperless

and fully automated procedures.³ The vision was that one could generate the information systems to support processes simply by modeling those processes. WFM systems attempt to realize this vision. DOMINO was the first European WFM system.⁵ Just like SCOOP and OfficeTalk, DOMINO used *Petri nets*—invented by Carl Adam Petri in the 1960s—to model processes. Several U.S.-based researchers picked up on the idea, working on both the theory and application of Petri nets. Interestingly, the pioneers in the field recognized

the importance of concurrency and, therefore, used Petri nets.

In the 1990s, many WFM products followed. Many people, including myself, expected that WFM systems would be used everywhere. However, despite a broad choice of products, adoption was limited. BPM systems^{1,7} extended WFM systems with more management and analysis capabilities, but these systems also suffered from limited adoption.

Between 1975 and 2005, the epicenter of process management research gradually shifted from the U.S.

Figure 1. Overview of process mining.


to Europe. Moreover, process modeling became more popular, and over time, the BPMN became the de facto standard.¹ Despite the widespread use of BPMN to model operational process models, the effect on improving processes was limited. One can argue that handmade BPMN models often have little to do with reality and are not taken seriously. Trying to implement these simplistic models using WFM/BPM technology is destined to fail. Therefore, these purely model-driven approaches became less popular over the last decade.

Data-Driven Process Management

Purely model-driven approaches tend to be disconnected from reality; process mining aims to address the problem. Organizations are not interested in handmade models that do not capture a constantly changing reality. I was one of the first to witness this, and at the end of 1990s, began a systematic, large-scale effort to learn the actual processes using event data. Unlike earlier approaches, which were either not data-driven or not process-centric, process mining combines event data and explicit process models.

Figure 1 provides an overview of process mining.⁸ First, event data is

extracted from information systems. These may be ERP systems, such as SAP and Oracle; CRM systems, such as Salesforce; or custom-made or domain-specific systems, such as healthcare information systems. Each event refers to an activity, has a timestamp, and may refer to any number of objects (orders, customers, items, and deliveries). It is often assumed that there is a special object, called the case, that relates and connects the individual events. Based on such event data, a process model can be automatically discovered. The model may describe all behavior or just mainstream behavior. It can also be adapted to describe what should have happened. Hence, process models may be descriptive or normative and are connected to event data using replay or alignment techniques.⁸ Through this connection, it is possible to diagnose performance and conformance problems. For example, the model can be annotated with information about deviations and bottleneck information. It is also possible to automatically conduct root-cause analysis and even predict performance and conformance problems.

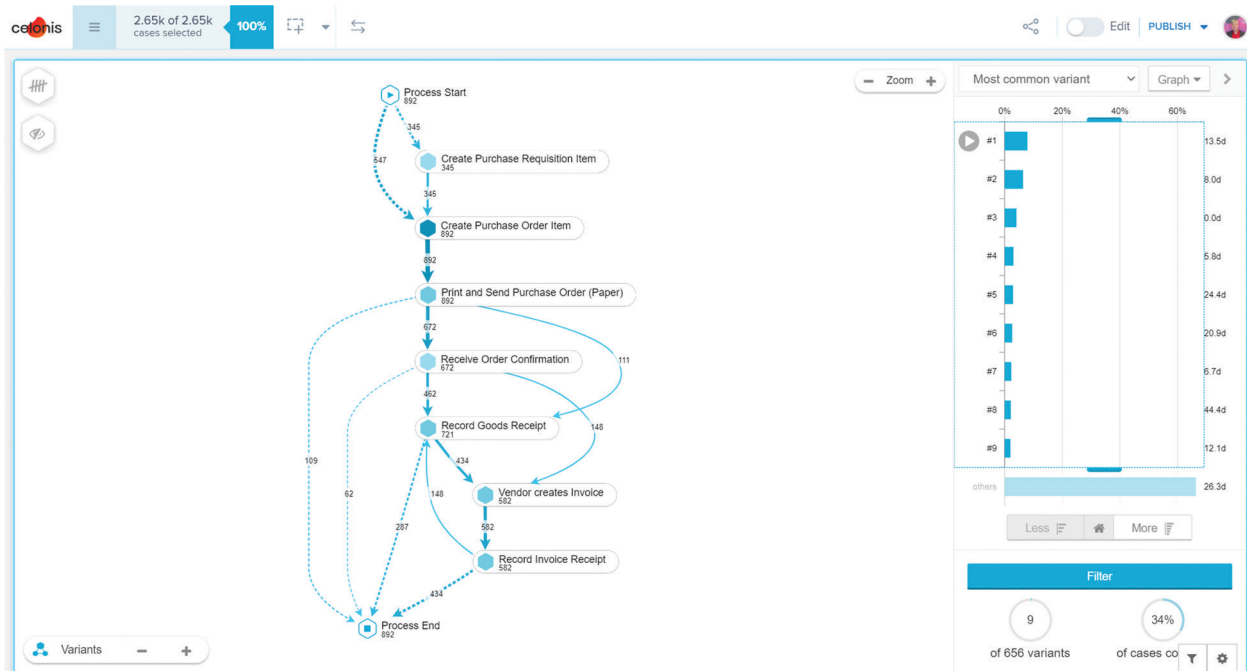
The *Alpha algorithm*, which I developed, was the first algorithm to discover concurrent processes from event

logs.⁸ Given a directly-follows complete event log generated by simulating a structured workflow net, the algorithm is guaranteed to produce a behaviorally equivalent Petri net. The two main approaches for conformance checking are *token-based replay* and computing *alignments* between log and model.⁸ In general, process mining-based algorithms tend to be very different from mainstream data-mining and machine-learning approaches.

Pain-Driven Automation

Process mining can be used to uncover so-called “execution gaps,” where reality is different from what is expected or desired. Conformance checking can be used to diagnose known “execution gaps,” and process discovery may reveal problems nobody was aware of. However, to improve the process, such “execution gaps” need to trigger improvement actions. Here automation again plays a role. As mentioned, traditional WFM/BPM technology was not very successful. However, one may use process mining to target the “execution gaps” in the process without trying to replace existing systems. Unnecessary rework or repetitive actions can be detected and bypassed or done automatically. Here, low-code integration platforms make it easy to trigger

Figure 2. Screenshot of the Celonis Execution Management System (EMS) while discovering a process model for a Purchase-to-Pay (P2P) process.



exactions in existing systems—for example, automatically sending an email message, redistributing work, or executing an SAP transaction.

Celonis Execution Management System

The Celonis Execution Management System (EMS) supports all process-mining capabilities shown in Figure 1. Celonis was founded in 2011 by Alexander Rinke, Bastian Nominacher, and Martin Klenk. They discovered the early papers on process mining and created a process-mining start-up that is widely considered one of Europe’s most successful young IT companies. The author is the chief scientist at Celonis and served since 2015 as its chief academic advisor.

Celonis has 1,800 employees and 2,000 customers. Some Celonis installations have more than 5,000 active users while handling more than 500 million cases and 10 billion events. The current valuation is 11.1 billion USD (after obtaining \$1 billion in Series D Round in June 2021). Celonis is clearly the market leader in process mining.

Figure 2 shows a screenshot of the Celonis EMS while discovering a process. Celonis was the first process-mining system not focused exclusive-

ly on data scientists, but on engaging the people who are actually involved in the processes. Because hundreds or even thousands of organizations analyze similar processes, it is possible to prepackage evidence-based knowledge. Celonis also supports action-oriented process mining with its embedded low-code integration platform, which can support hundreds of applications, including SAP, Oracle, and Salesforce. This way, process insights can be turned into improvement actions.

Conclusion

Process mining can be viewed as “taking X-rays” to uncover known and unknown “execution gaps,” while automation can be seen as the surgery undertaken to address those gaps and improve the process in a targeted way. Classical WFM/BPM approaches can be seen as surgery that is performed before taking X-rays to diagnose the actual problems. The Celonis EMS aims to combine both—that is, using process mining to uncover, diagnose, and predict performance and compliance problems. These problems are subsequently addressed through targeted forms of automation.

Although most process-mining research is conducted in Europe,

process mining is applied globally, and adoption is rapidly growing in the U.S. and Asia-Pacific. For example, more than 50% of the Fortune 500 companies are exploring process mining. Yet, we are just at the beginning, with many opportunities for cutting-edge research and innovative industrial applications. 

References

1. Dumas, M., La Rosa, M., Mendling, J., and Reijers, H. *Fundamentals of Business Process Management*. Springer (2018).
2. Ellis, C. and Nutt, G. Computer science and office information systems. *Computer Surveys* 12, 1 (1980), 26–60.
3. Giuliano, V. The office of the future, *Business Week* (June 30, 1975), 48–70.
4. Kerremans, M., Srivastava, T., and Choudhary, F. Market guide for process mining. Gartner Research Note G00353970 (2021), <https://www.gartner.com/en/documents/3991229/market-guide-for-process-mining>.
5. Kreifelts, T. DOMINO: Ein system zur abwicklung arbeitsteiliger vorgänge im büro. *Angewandte Informatik* 26, 4 (1984), 137–146.
6. Reinkemeyer, L. *Process Mining in Action: Principles, Use Cases and Outlook*. Springer, Berlin (2020).
7. van der Aalst, W. Business process management: A comprehensive survey. *ISRN Software Engineering*, 507984 (2013), 1–37.
8. van der Aalst, W. *Process Mining: Data Science in Action*. Springer, Berlin (2016).
9. Vogelgesang, T. et al. Celonis PQL: A query language for process mining. In *Process Querying Methods*, Springer (2021).
10. Zisman, M.D. Representation, specification and automation of office procedures. Ph.D. thesis, University of Pennsylvania, Wharton School of Business (1977).

Wil van der Aalst is a professor at RWTH Aachen University, München, Germany.

Copyright held by author/owner. Publications rights licensed to ACM.

BY SHAUKAT ALI, TAO YUE, AND RUI ABREU

When Software Engineering Meets Quantum Computing

OVER THE LAST few decades, quantum computing (QC) has intrigued scientists, engineers, and the public across the globe. Quantum computers use quantum superposition to perform many computations, in parallel, that are not possible with classical computers, resulting in tremendous computational power.⁷ By exploiting such power, QC and quantum software enable many applications that are typically out of the reach of classical computing, such as drug discovery and faster artificial intelligence (AI) techniques.

Quantum computers are currently being developed with a variety of technologies, such as superconducting and ion trapping. Private companies, such as Google and IBM, are building their own quantum computers, while public entities are investing in quantum technologies. For example, the European Union Commission is spending €1 billion on quantum technologies (“EU’s Quantum Flagship Project’s Website”^a). Currently, the key goal for quantum

computers is to reduce hardware errors that limit their practical uses. Regardless of the eventual technology that wins the quantum hardware race, the key enabler for building QC applications is quantum software (see Figure 1).

Quantum software needs to be supported with a quantum software stack, ranging from operating systems to compilers and programming languages, (see examples in Table 1) as postulated by Bertels et al. from the University of Porto.³ *Quantum software engineering (QSE)* enables the cost-effective and scalable development of dependable quantum software to build revolutionary quantum software applications in many domains—for example, finance, chemistry, healthcare, and agriculture (see Figure 1 and Table 1). However, effective quantum software applications cannot be developed with classical software engineering methods due to quantum computing’s inherent characteristics—for instance, superposition and entanglement. Thus, we need to build novel QSE methodologies (with tool support) that cover different phases of QSE, possibly including requirements engineering, modeling, coding, testing, and debugging as shown in Figure 1.

In this article, we first present a general view of quantum computing’s potential impact, followed by some highlights of EU-level QC initiatives. We then argue the need for QSE, present the state of the art of QSE from multiple aspects (testing, for example) by comparing quantum computers with their classical counterparts, and shed light on possible research directions.

The Impact of Quantum Computing

Quantum computing is primed to solve a broad spectrum of computationally expensive societal and industrial problems. Notable examples include accelerated drug discovery and vaccine development in healthcare,

^a <https://qt.eu/about-quantum-flagship/introduction-to-the-quantum-flagship/>

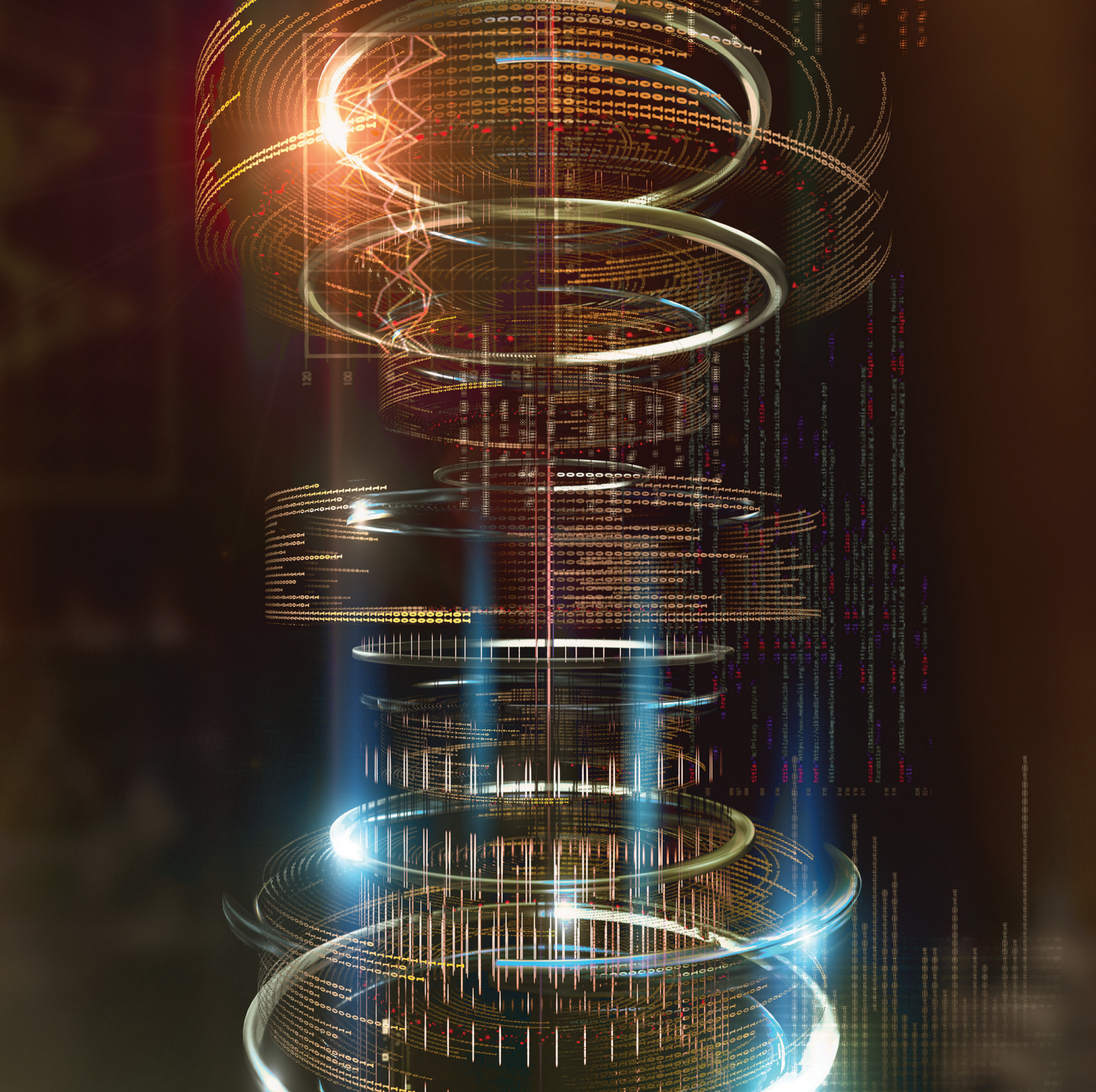


IMAGE BY CARLOS CASTILLA

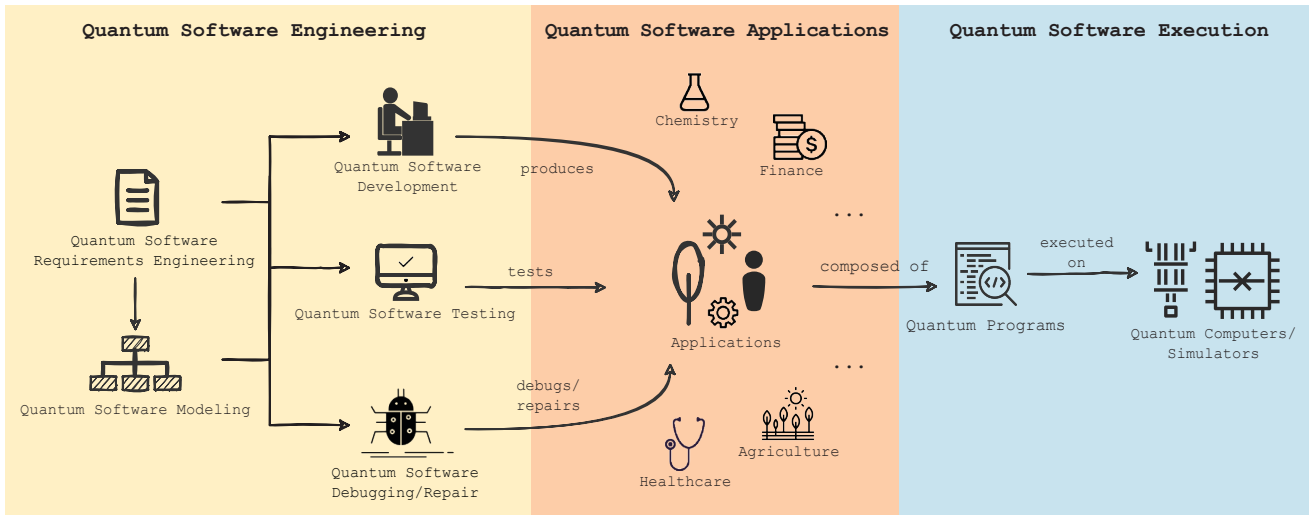
portfolio management and optimization in finance, and complex simulations in physics to better understand our universe. As a result, QC's success will inevitably and significantly impact our day-to-day lives and revolutionize most industries across many domains. Such impact must be realized via quantum software, the development of which should be systematically powered by QSE. Scientifically speaking, QSE will open new areas of research to develop real applications by fostering

research communities across disciplines (such as computer science, software engineering, mathematics, and physics) and interactions with other fields such as medicine, chemistry, and finance. Table 1 summarizes various dimensions of QC with examples.

EU-level quantum initiatives. Efforts to build quantum computers in Europe are increasing. VTT-Technical Research Centre of Finland, together with IQM, aims to build Finland's first 25-qubit, fully functional quan-

tum computer by 2024. In Sweden, the Wallenberg Centre for Quantum Technology at Chalmers is building a superconducting quantum computer capable of up to 100 qubits. The Future and Emerging Technologies' Quantum Technologies Flagship program also funds projects to build quantum computers. For example, AQTION is building Europe's first ion-trapped quantum computer, while OpenSuperQ is focused on building a 100-qubit superconducting QC. To boost research

Figure 1. An overview of quantum software engineering.



on the development of novel QC applications in Germany, Fraunhofer installed an IBM Quantum System One to provide access to organizations interested in developing QC applications. Finally, NordiQuEst is a new collaborative effort between four Nordic countries and Estonia to build a dedicated Nordic-Estonian QC ecosystem that will integrate various quantum computers and emulators and make them accessible to the Nordic-Estonian region to accelerate QC research, development, and education.

Why Quantum Software Engineering?

Building practical and real-life QC applications requires the implementation of quantum algorithms as software. Learning from the classical computing realm, developing dependable software entails following a *software development life cycle (SDLC)*, which typically includes requirements engineering, architecture and design, development, testing, debugging, and maintenance phases.

Given that quantum software de-

velopment is relatively new, an SDLC for quantum software doesn't exist. However, quantum programming languages are available to implement quantum algorithms (see examples in Table 1). In their current state, these languages allow programming at the lower level—for instance, as quantum circuits consisting of quantum gates. Figure 2 shows a quantum program example in IBM's Qiskit performing quantum entanglement, its equivalent quantum circuit in the middle, and execution result on the right side.

Programming quantum circuits is challenging, as evidenced in the example, because it requires a specialized background in quantum physics, including an understanding of how quantum gates work. Unfortunately, classical computing programmers do not often possess such a background, thus making it difficult for them to program quantum computers. Moreover, in the context of quantum SDLC, quantum programming is just one aspect; attention must be given to other SDLC phases, such as requirements, design, and architecture; verification and validation; and maintenance.

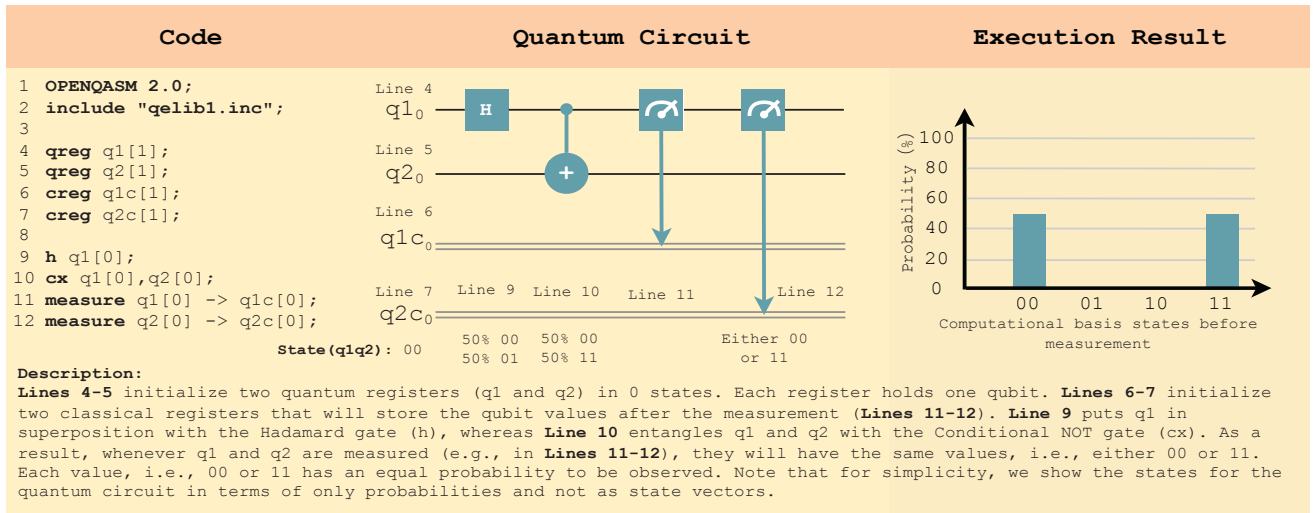
Classical vs. Quantum Software Engineering

Quantum software requirements engineering and quantum software modeling. Due to the increasing complexity of software application domains, requirements engineering is critical, as it is the process of eliciting

Table 1. Various dimensions of quantum computing with examples.

Dimension	Examples
Applications	Radiotherapy optimization, speeding up AI algorithms, empowering modeling and simulations in aerospace and physics, drug discovery, vaccine development, cryptography, portfolio management, among others.
QC programming languages	OpenQL by TU Delft Netherlands, Silq by ETH Zürich Switzerland, Q# by Microsoft, Qiskit by IBM, Cirq by Google
QC platforms	Quantum Inspire from QuTech Netherlands, Microsoft Quantum computing platform, and IBM Quantum Experience
Open source software	<i>Quantum compilers:</i> BQSKit, D-Wave's qbsolv <i>Computer simulators:</i> QuEST, QuPy <i>Editors:</i> ProjectQ, QisKit Circuit Composer
Industry	Norwegian Quantum Computing Group, IQM, Cambridge Computing, Qbee.eu, AegIQ, Algorithmiq, QBaltic, Arqit
Events	Quantum Software Engineering Workshops co-located with ICSE, QC Talks at University of Porto, Portugal; Quantum Software Engineering and Technology Workshop; International Workshop on the QuANTum SoftWare Engineering & pRogramming

Figure 2. Quantum entanglement program in OpenQASM together with its quantum circuit and execution result.



ing, specifying/modeling, managing requirements, among others. During the process, various stakeholders—domain experts and software developers, for example—interact, and the resulting requirements serve as key artifacts to drive software design and development. From this perspective, we consider that requirements engineering in the quantum world aligns with requirements engineering of its classical computing counterpart (see Table 2 for details). However, QC brings new challenges.

First, its software engineers often find its application domains—for instance, radiotherapy optimization or drug discovery—hard to comprehend. Second, quantum software engineers must also equip themselves with basic knowledge about quantum mechanics, linear algebra, algorithms and their analysis, and more. Therefore, requirements engineering is very important for easing communication among various stakeholders while raising the level of abstraction in understanding the domain and linking the domain to analysis, design, and implementation. To the best of our knowledge, requirements engineering for quantum software is an uncharted area of research. There has not been any publication in this area yet. We argue that, as with classical software, quantum software engineering requires the development of novel elicitation, specification, modeling, analysis, and verification methods.

Table 2. Comparison of classical and quantum computing.

Aspects	Classical	Quantum
Information encoding	Bits	Qubits
Building block	Classical gates, that is, binary operations on binary inputs, such as NOT (which negates an input bit, such as, 0 to 1 and 1 to 0).	Quantum gates perform operations on qubits. For example, the <i>Hadamard</i> gate puts qubits in superposition (see Figure 2). A quantum circuit is formed with a set of gates (see Figure 2).
Logic function	Operates on a register of n bits.	Operates on a register of n qubits. Special Characteristic: Reversibility.
State	0 or 1 for one bit For example, two bits hold exactly <i>one</i> value at a time from the following possible values 00, 01, 10, 11.	Superposition: Quantum program can exist in multiple states at the same time. Entanglement: Two entangled qubits or registers exist in a single quantum state. For example, two qubits can hold four values at once: 00, 01, 10, and 11.
Programming process	Varies when using different programming languages, spanning from low-level assembly languages to high-level languages, for example, Python.	The current practice is at the assembly language (for example, with Open Quantum Assembly Language) or designing quantum circuits.
Testing and debugging	(1) Reading intermediate states possible in some cases; (2) Some test oracles are probabilistic; (3) Fewer hardware errors; and (4) Direct breakpoints possible for interactive debugging.	(1) Directly reading intermediate states destroys superposition; (2) Most test oracles are probabilistic; (3) Many hardware errors; (4) Direct breakpoints not possible; and (5) Facing the decoherence problem.
Software development	(1) Many SDLC, such as waterfall and agile; (2) Lots of frameworks for transferring high-level designs into low-level implementations; and (3) Intuitive programming languages compared to quantum programming languages.	(1) No well-established SDLC; (2) Lack of abstraction, for example, non-existent mechanisms for translating high-level designs to gate-level implementations; and (3) Less intuitive programming languages.

Quantum finite-state machines, and the study of their formal properties, have been investigated in the literature.⁸ However, their application to quantum software modeling remains unstudied. Recently, there has been an increasing interest in extending the Unified Modeling Language (UML) to model quantum software, mainly in the classical software engineering community as highlighted by European researchers.² More research is needed, though, to determine whether extending UML is sufficient or more domain-specific modeling solutions are required. In general, there are many opportunities for quantum software modeling, such as developing novel and intuitive quantum modeling notations and methodologies, verification and validation with quantum software models, and empowering code/circuit generation.

Quantum software testing. It is important to ensure quantum programs are correct—that is, they can deliver their intended functionalities. Testing quantum programs is difficult compared to classical software due to their inherent characteristics, including their probabilistic nature; computations in superpositions; the use of advanced features, such as entanglement; a difficulty in reading or estimating quantum program states in superposition; and a lack of precise test oracles. Thus, there is a need for novel, automated, and systematic methods for testing quantum programs. Quantum software testing is garnering increased attention, and several papers have recently been published, with significant contributions from European researchers.^{1,6,9,10}

Several research areas need to be explored, such as how to define and check (with relevant statistics, for instance) quantum test oracles without destroying superposition, and how to cost-effectively find test data that can break a quantum program. Given hardware noises in quantum computers, testing techniques must also be noise-aware. In general, we foresee the need to build theoretical foundations of quantum software testing, including coverage criteria, test models, and test strategies. Test strategies consist of test oracles, test data, and test cases, and can be designed by consid-

ering fault types, metamorphic testing to deal with test oracle issues, and mutation analysis. To maximize the benefit, all test techniques are expected to be independent of a quantum programming language. Furthermore, we need practical applications, extensive empirical evaluations of testing techniques, and the creation of benchmarks for the community. Several automated quantum software testing tools have recently been developed, with major contributions by European researchers such as Quito, Muskit, QuSBT, QsharpCheck, and QuCAT.

Quantum software debugging.


Observed failures in quantum programs—for instance, found with testing—need to be diagnosed with the debugging process to isolate and patch the code to fix the failure. This process typically comprises multiple tactics usually found in debuggers, such as relying on print statements in code to achieve interactive debugging. Similarly, we need quantum software debugging tactics, implemented in debuggers, to cost-effectively diagnose and resolve quantum software failures. However, the development of effective debugging techniques faces several challenges as discussed in a key debugging work:⁴ an inability to directly monitor quantum software states in superposition; the understanding of quantum software states, when possible (for instance, in quantum computer simulators), can be unintuitive; and a lack of best practices, in general, to perform debugging.

Several research opportunities exist for debugging quantum programs: tailoring classical debugging tactics (backtracking and cause elimination, for example) to debug programs on quantum simulators and developing novel tactics to debug on real quantum computers; novel visualization approaches to inspect values without the need to measure quantum states, with intuitive visualizations comprehensible by humans; and novel ways to infer quantum software states using statistical⁴ and projection-based assertions (for example, see Li et al.⁵), in addition to developing novel assertion types.

Conclusion

Quantum computing is on the rise

and will, no doubt, revolutionize many technologies. It will transform our understanding of and the way we deal with complex problems and challenges. Quantum software engineering is key to the systematic and cost-effective creation of tomorrow's powerful, reliable, and practical QC applications.

Compared with classical computing, QC's inherent complexity and its complex application domains—drug discovery, for example—present new multidimensional challenges that emphasize the significance of QSE. Fascinated by this observation, we presented in this article the key highlights of QC activities in Europe, key QSE innovations (when compared with classical software engineering), and open QSE research directions. This is the time to embrace QC and form the QSE community in Europe and globally. 

References

1. Ali, S., Arcaini, P., Wang, X., and Yue, T. Assessing the effectiveness of input and output coverage criteria for testing quantum programs. In *Proceedings of the 14th IEEE Conf. on Software Testing, Verification, and Validation* (2021).
2. Ali, S. and Yue, T. Modeling quantum programs: Challenges, initial results, and research directions. In *Proceedings of the 1st ACM SIGSOFT Intern. Workshop on Architectures and Paradigms for Engineering Quantum Software*. (2020).
3. Bertels, K. et al. Quantum accelerator stack: A research roadmap. (2021), arXiv:2102.02035.
4. Huang, Y. and Martonosi, M. Statistical assertions for validating patterns and finding bugs in quantum programs. In *Proceedings of the 46th Intern. Symp. on Computer Architecture* (2019).
5. Li, G. et al. Projection-based runtime assertions for testing and debugging quantum programs. In *Proceedings of the ACM on Programming Languages* 4 (2020); doi:10.1145/3428218.
6. Mendiluze, E., Ali, S., Arcaini, P., and Yue, T. Muskit: A mutation analysis tool for quantum software testing. In *Proceedings of the 36th Intern. Conf. on Automated Software Engineering—Tool Demonstrations track* (2021).
7. Mermin, N.D. *Quantum Computer Science: An Introduction*, Cambridge University Press (2007).
8. Tian, Y., Feng, T., Luo, M., Zheng, S., and Zhou, X. Experimental demonstration of quantum finite automaton. *npj Quantum Information* 5, 1 (2019), 56; doi:10.1038/s41534-019-0163-x
9. Wang, X., Arcaini, P., Yue, T., and Ali, S. Application of combinatorial testing to quantum programs. In *Proceedings of the IEEE Intern. Conf. on Software Quality, Reliability, and Security* (2021).
10. Wang, X., Arcaini, P., Yue, T., and Ali, S. Generating failing test suites for quantum programs with search. In *Proceedings of the 13th Symp. Search-Based Software Engineering* (2021).

Shaukat Ali is a chief research scientist, research professor, and head of department at Simula Research Laboratory, Oslo, Norway.

Tao Yue is an adjunct research scientist at Simula Research Laboratory, Oslo, Norway.

Rui Abreu is a professor at the University of Porto, Portugal.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. <http://creativecommons.org/licenses/by/4.0/>