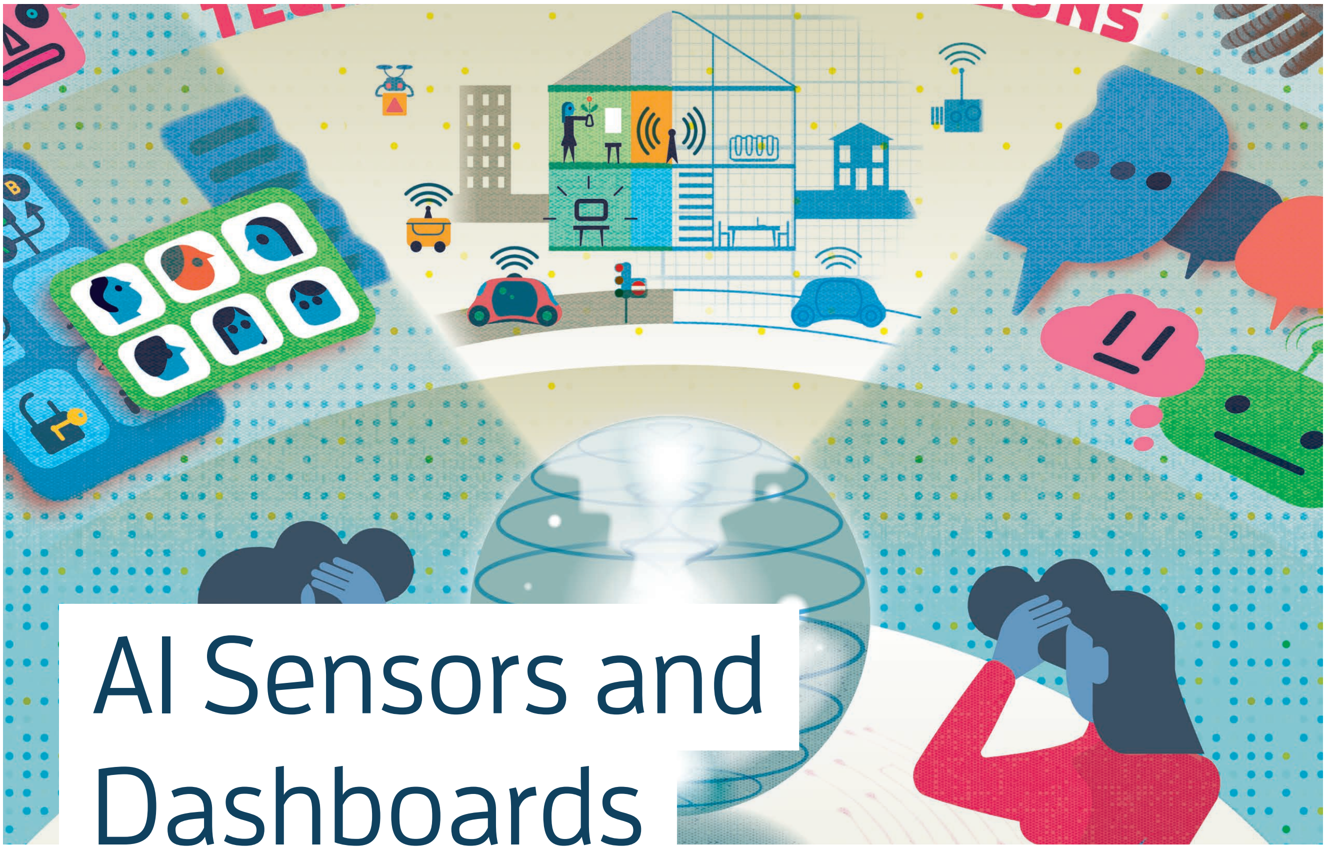

AI Sensors and Dashboards

Huber Flores, University of Tartu

Measurements are fundamental to our understanding and control of technology. We predict AI sensors and dashboards that monitor AI inference capabilities and its performance and enable users to interact with AI, promoting responsible usage, building trust, and ensuring compliance with ethical and regulatory standards.

The adoption of artificial intelligence (AI) in our society is imminent. Despite its enormous economic impact, the lack of human-perceived control and safety is redefining the way in which emerging AI-based technologies are developed and deployed in systems and end applications. New regulatory requirements to make AI trustworthy and responsible are transforming the role that humans play when interacting with AI, and consequently, AI is now not just creating new opportunities and markets, but it is doing it while preserving the fundamental rights and liberties of individuals. In this article, AI sensors and dashboards are predicted to become an integral part of AI solutions. AI sensors can gauge the inference capabilities of the technology, whereas AI dashboards can allow individuals to monitor and tune it transparently.

AI TRUSTWORTHINESS

The AI market value is expected to increase from US$100 billion to US$2 trillion by 2030 according to reports from Statista and numerous other sources.[1] This exponential growth emphasizes the imminent adoption of AI in everyday applications. AI's disruptive inference process

Digital Object Identifier 10.1109/MC.2024.3394056
Date of current version: 26 July 2024

has baffled the world as an increased number of users reported and perceived human-like reasoning when interacting with powerful AI-based models available online,[2] for example, ChatGPT, Ernie, and Gemini. This advanced performance seemed incomprehensible at first hand, leading to the release of an open global petition in March 2023 for slowing down AI developments for at least six months.[3] Indeed, the opacity and black-box characteristics in machine and deep learning models have demonstrated high inference capabilities when trained at scale, but since its internal mechanics are obfuscated and unclear, the use of AI models fostered distrust and unsafety for human operators and developers.[3] Current development practices that ensure the trustworthiness of software, for example, formal verification, are not applicable to the construction of AI models.[4] Thus, new methods for gauging and controlling the capabilities of AI are key to making the technology trustworthy and fostering responsible deployments of AI in everyday applications and interactions with humans.

All economic and regulatory systems worldwide recognize the need to cultivate trustworthiness in digital technologies, and AI is the key one to focus on. The lack of transparency, accountability, and resilience in emerging AI-based technologies is a global concern, which has led to the imposition of strict regulations for their development. National and international sovereignty over AI-based applications and services aims to ensure public trust in AI usage. As a result, the European Union (EU) strategic plan for AI adoption, outlined in the EU General Data Protection Regulation 2016/679 and EU AI Act,[5] has emerged and become an international benchmark since the early stages of AI developments. Likewise, the United States has acknowledged the significance of regulating AI usage through its U.S. AI Act Executive Order 13859/13960.[6] China has also emphasized the importance of regulating generative AI developments as crucial steps in developing trustworthy AI technology.[7]

AI's inference capabilities and its performance can be characterized through the use of different trustworthy properties. AI trustworthiness is defined by extending the properties of trustworthy computing software with new considerations that take into account the probabilistic and opaque nature of AI algorithms and quality of training data.[8] Trustworthy AI is valid, reliable, safe, fair, free of biases, secure, robust, resilient, privacy preserving, accountable, transparent, explainable, and interpretable.[4] Notice, however, that AI trustworthiness is an ongoing process whose definition is evolving continuously and that involves collaboration among technologists, developers, scientists, policymakers, ethicists, and other stakeholders. Moreover, the mapping and implications of the ethical and legal requirements of technical solutions remain unclear.

In this article, we predict *AI sensors and dashboards* as a research vision that is an integral part of the adoption of AI and its interactions with individuals. An AI sensor can aid in monitoring a specific property of trustworthiness, whereas an AI dashboard can provide visual insights that allow humans to gauge and control the inherent properties of AI based on human feedback. Moreover, it has been demonstrated that trustworthy properties can be considered tradeoffs when implemented in practice,[9,10] suggesting that modifying one property can impact others, for example, robustness versus privacy, accuracy versus fairness, and transparency versus security. Thus, AI sensors are envisioned to interact and establish negotiations between them to obtain a balanced level of trust based on the type of application at hand.[11] Our prediction is that all applications and systems implementing AI-based functionality will provide a dashboard and will be instrumented with sensors that measure, adjust, and guarantee trustworthiness, such that individuals interacting with AI can be aware of its trust level. We highlight the technical challenges, current technological enablers to build upon, and implications of realizing this vision.

CONCEPTUAL BACKGROUND

The responsible deployment of AI in everyday applications is key to scaling up the adoption of the technology. To analyze this, we first reflect on current AI regulations and their implications for software development practices. After this, we then highlight existing solutions aimed at characterizing the inference process of AI. With this information, we introduce the concept of AI sensors and dashboards.

Control over AI via regulations

Regulations over AI seek to promote the responsible development and deployment of AI technologies. Europe has crafted an extensive and comprehensive legislative proposal that highlights the possible risks and unwanted practices for the development of AI models. Moreover, it also emphasizes the assessment of AI-based technologies to verify transparency and adherence to human rights as a way to foster trust in society.[5] To fulfill these goals, regulations provide guidelines and compliance support for handling data and developing software architectures.

Consequently, software engineers and other practitioners must consider new requirements, such as data traceability, minimization, rectification, and erasure. They also address system security, privacy, and risk management. Similar and overlapping principles are also described in the U.S. AI Act,[6] China's regulations over generative AI,[7] and those of other countries, like Japan, Brazil, and Canada.

Modern applications and AI

Modern applications have evolved significantly beyond classical client-server architectures. Currently, modern architectures incorporate machine and deep learning pipelines (AI components) that collect data from user interactions and exploit them to train AI models—using either centralized or distributed approaches.[12] In practice, analyzing the inference capabilities of AI thus involves evaluating: 1) the trained AI model itself, 2) the training data, and 3) the overall AI pipeline that constructed the model. However, modern applications with integrated AI lack features to monitor the inference capabilities of AI effectively. As a result, they fall short of complying with AI regulations. Ongoing efforts to communicate the internal logic of AI models have led to the development of monitoring solutions, where the performance characteristics of AI models can be quantified and visualized in terms of metrics, such as accuracy and F1-score. Examples of this include TensorLeap (https://tensorleap.ai/), Neptune AI (https://neptune.ai/), and Comet ML (https://www.comet.com/site/). Advanced monitoring tools that facilitate the comprehensive characterization of AI trustworthiness are a promising approach to engaging humans in the tuning of AI as well as verifying its internal inference behavior.

Toward AI sensors and dashboards

Sensors are commonly instrumented within applications to enable the monitoring of their functionality during runtime. Sensors are fundamental mechanisms for data collection and measurements. AI sensors are envisioned as software-based mechanisms, for example, virtual sensors.[13] A virtual sensor thus is a program that characterizes or continuously profiles the behavior of certain implemented functionality. Since AI models are updated on time (retrained as new data are obtained), AI sensors observe how these changes influence different characteristics of the models, for example, resilience, accuracy, and fairness to mention a few. AI sensors can also potentially learn from these observations to determine when models have been alternated drastically by contributions, for example, possible attacks. In turn, an AI dashboard communicates through visual insights the measurements collected by the AI sensors, such that individuals can inspect, assess, and tune the behavior of AI.

ENABLING TECHNOLOGIES

AI sensors and dashboards simplify the complexity of advancing the monitoring tools of AI trustworthiness. Building these tools, however, requires building upon existing technologies. Thus, we continue by describing the technological enablers supporting the implementation of AI sensors and dashboards in practice.

Path to AI sensors

AI sensors are envisioned to be instrumented within modern applications at the code level, such that it is possible to analyze the (serialized) AI model (in JSON/YAML), the dataset, and its respective pipeline. Functioning as application programming interfaces (APIs), AI sensors leverage standard technologies for system integration and interoperability. AI sensors are designed with a clear separation between their interface (client API) and functionality (deployed in a back end), ensuring lightweight instrumentation routines and reducing processing costs in end applications. At the same time, this clear separation allows for changing the functionality of the AI sensors without modifying the end application. This is useful as currently there is a mismatch between technical and legal trustworthiness. Upgrading the functionality of an AI sensor can then become simple by adopting system architecture patterns like microservices.

In addition, another important reason to separate interface and functionality is that several AI sensors are required to be instrumented within an application, such that it is possible to characterize different trustworthy properties. This can cause the processing requirements of applications to become higher. Thus, outsourcing the functionality to remote infrastructure can be helpful to avoid introducing extra processing overhead in applications. Furthermore, AI sensors instrumented in an application are meant to interact between them, such that the autonomous tuning of trustworthiness can be achieved based on the type of application or context at hand. This autonomous tuning (or negotiations) also requires further processing capabilities that allow AI sensors to reach an agreement regarding the level of trust to be provisioned to users. This is particularly helpful in dynamic situations where the use of data becomes context dependent,[14] requiring, in some cases, consent from surrounding individuals to use their data. In such cases, AI sensors can act on behalf of users to aid in automatizing the data process of data handling and management. Notice, however, that users are required to be aware of their preferences and how these are configured within applications.

Path to AI dashboards

An AI dashboard communicates through concise visual insights the measurements collected by the AI sensors, such that individuals can inspect, assess, and tune the behavior of AI. Notice that while the quantified information of all trustworthy properties can be presented, the type of application from which trustworthiness is estimated can play a role in presenting the results in the AI dashboard. As an example, fairness can be an important factor for employment-, healthcare-, and finance-related applications, but it may be of less importance for autonomous applications like self-driving cars and drone delivery. This suggests that visualization through an AI dashboard depends on the type of application, requiring methods to reorganize content, such as hierarchy analysis or progressive disclosure mechanisms.[15] Once information is available in the AI dashboard, tuning or providing feedback to enhance AI inference capabilities is not an individualized process but requires specific stakeholders, such as domain- or application-specific experts to adjust AI models based on user insights.

AI dashboards facilitate model tuning for experts and provide insights into inference capabilities for all users. For example, in an AI model for bank loans, end users can assess the fairness of the model through the dashboard, but only designated expert stakeholders can apply user feedback to refine the model. Tuning of AI models can be achieved through several existing open source and proprietary tools and libraries, including Ray Tune (https://ray.io/), Optuna (https://optuna.org/), Hyperopt (https://hyperopt.github.io), Vizer (https://github.com/vizier-db), Microsoft NNI (Neural Network Intelligence, https://nni.readthedocs.io), Keras Tuner (https://keras-team.github.io/keras-tuner/), and SigOpt (https://sigopt.com/). Naturally, model tuning may compromise AI developments, requiring the use of secure technologies to ensure that AI models are not hampered intentionally.

IMPACT

AI sensors and dashboards are predicted to be introduced in applications, as shown in Figure 1. We next highlight how AI sensors and dashboards can improve the perception and interaction of users with different types of applications.

Existing real-world applications

Currently, online applications already implement AI models to some extent, in the form of either recommendations or personal guidance for individuals. These applications request that users enable their history interactions with applications to improve their recommendation logic, providing better suggestions that match users' interests. Several existing applications provide coarse-grained estimates about this interest-matching characterization; for example, Netflix provides a matching score for movie recommendations. AI sensors and dashboards can provide additional benefits for these applications, providing fine-grained details on the considerations taken to reach this matching decision. As an example, consider an online bookstore (like Amazon); book recommendations are provided to users, but the details on how a recommendation is triggered are speculative to the users receiving them. AI dashboards can help users explore whether recommendations provided by the website were taken given different parameters, like demographic groups, age, type of behavioral interactions, and overall, a large variety of human patterns. AI sensors can provide additional fine-grained information regarding the model characteristics, such as privacy and biases, demonstrating that even simpler applications can rely on AI sensors and dashboards to improve the awareness of AI to individuals.

Autonomous applications

Thanks to the emergence of robust AI models for navigation and localization, autonomous technologies (like autonomous cars and drones) are now fully operational and deployed in urban areas, for example, delivery drones and autonomous cars.[16] The accountability of these technologies when facing unexpected crashes and abnormal behaviors remains a key challenge for their safe adoption.[17] Besides this, the lack of visual human operators causes distrust in users. AI dashboards running on the personal devices of users can potentially retrieve general information about AI in cars and drones, such that users can decide whether to use it or not. This information can include safety and performance trustworthiness metrics, highlighting the effective operations of the autonomous decision models. These dashboards can also provide and collect feedback over time from other users, increasing the usability and comfort of the technologies.

Personalized applications

Federated learning as a service has been proposed to build personalized applications for personal devices.[12] These applications train robust AI models over time in a collaborative manner as users encounter other individuals with similar preferences and interests. Since not all the updates to AI models are beneficial,[4] AI dashboards can provide insights on whether aggregation is beneficial or detrimental to personalized model performance. For instance, it may be that the data contributions and features are irrelevant for certain users. As a result, users can proactively decide whether to accept or reject certain contributions from others through the AI dashboard.

Metaverse applications

Augmented reality/virtual reality technologies exploit AI to provide advanced immersive experiences to users.[18] Indeed, generative AI can easily construct a large variety of different digital environments for users to experience. However, this adaptive functionality can hamper other functionality in the digital environment. For instance, the behavior of AI models in other objects can change significantly, reducing their robustness levels. Thus, AI sensors can then characterize and monitor over time the resilience and robustness of these objects when facing different environments. The AI dashboard can then provide this information to users to determine the level of operational

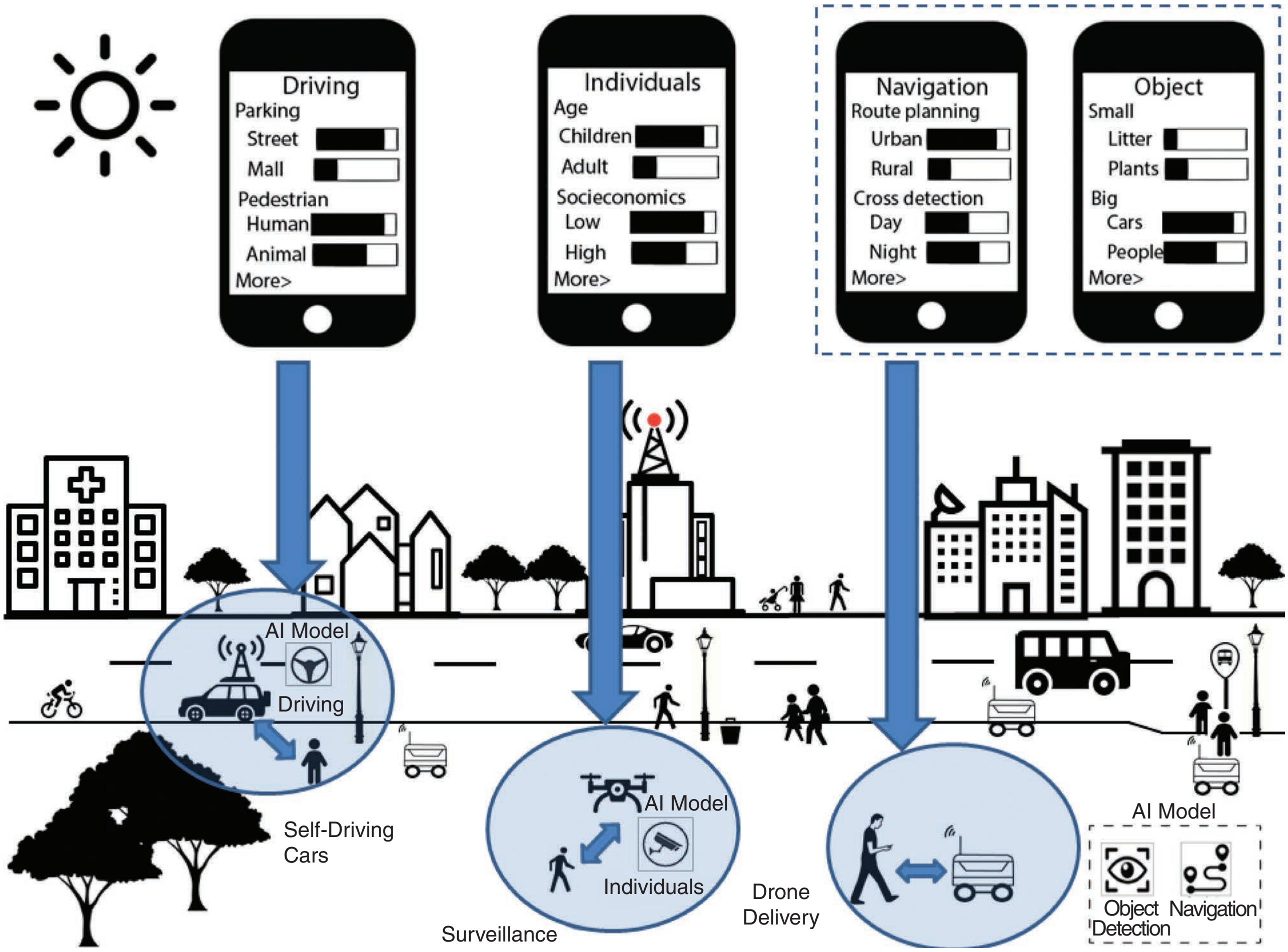

FIGURE 1. A vision of AI sensors and dashboards for modern applications.

immersive experience that a particular digital environment can provide without failures. AI dashboards can be presented to users as a part of their immersive experience and description of their virtual environment.

Generative applications

Generative data produced by AI models is key for augmenting and enriching scarce datasets.[19] This incidentally can influence the explainability and interpretability of models. Synthetic generated data can introduce biases in model inference. AI sensors can monitor the performance of models and their relationship with generated data. Potentially, AI sensors can adjust and balance the difference between real and synthetic data. Likewise, the AI dashboard can provide detailed information about how reliable the model is based on real measurements and provide insights about the amount of generative data supporting the AI model.

CHALLENGES AND FORESEEN DEVELOPMENTS

We next reflect on the current state of existing technologies and highlight the core challenges to overcome to achieve our vision.

Sensor instrumentation

By default, common practices for analyzing AI models are performed using a post-de facto verification approach.[8] This means that the AI model is analyzed once it is fully constructed, deployed, and functional. AI models can be instrumented with AI sensors using standard API routines. However, this is not a trivial task. As shown in Figure 2, building an AI model involves multiple steps abstracted into a pipeline.

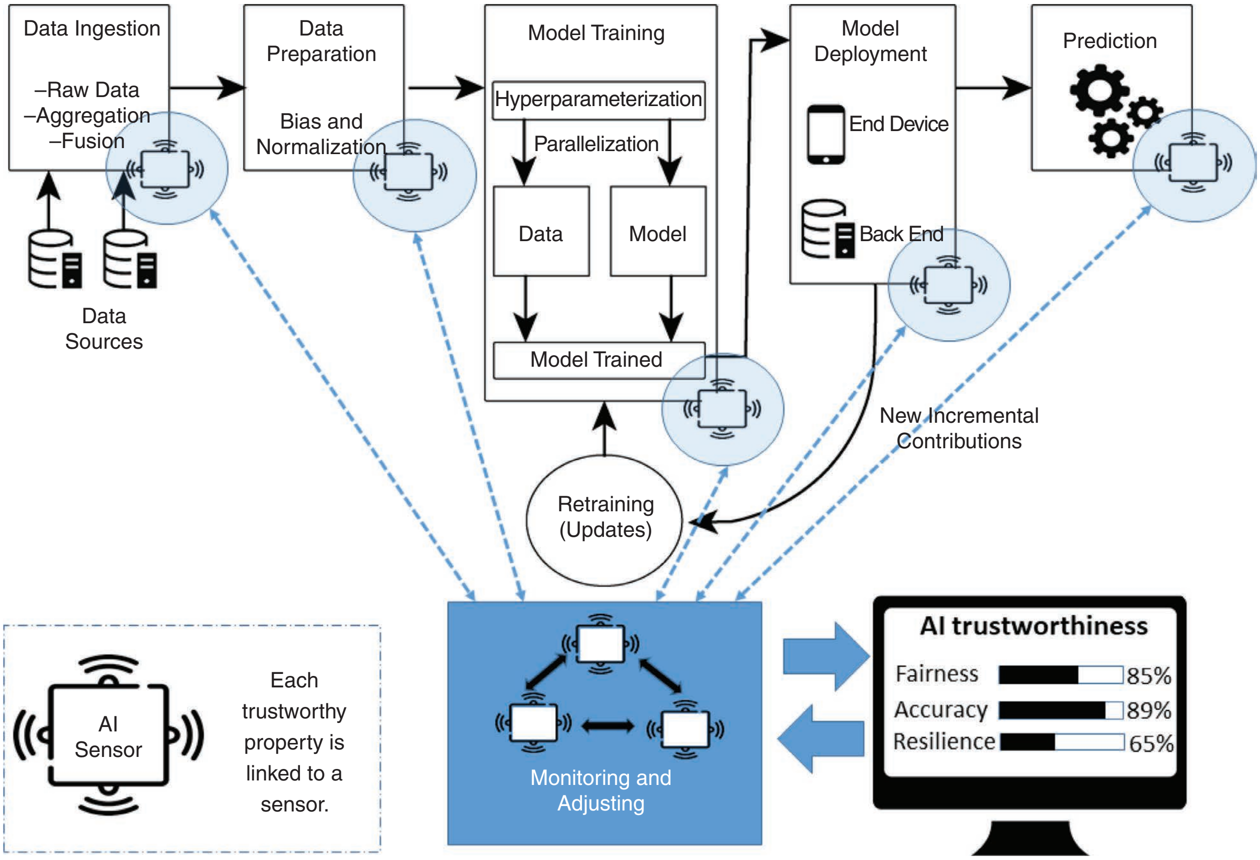

FIGURE 2. A standard machine learning pipeline instrumented with AI sensors and collecting measurements displayed in an AI dashboard.

Each step influences the overall resulting model that is produced, suggesting that the overall pipeline requires the instrumentation of AI sensors. For instance, it is possible to establish the level of fairness of a model before its construction just by analyzing its raw data, for example, using statistical parity or a data imbalance method, such as resampling.[9] Similarly, fairness can be derived once the model is fully operational or after each update, for example, using equal opportunity or equalized odds metrics.[9] Thus, a key challenge to enable AI sensors is to develop sensors tailored to monitor each step of the AI pipeline.

This has two implications

1. A trustworthy-by-design approach must be encouraged instead of a post-de facto analysis.
2. A single sensor for monitoring a specific trustworthy property may not be enough, requiring instead having multiple AI sensors of the same type embedded at different steps of the pipeline.

Another challenge is to develop loose instrumentation principles, such that AI sensors can be easily equipped into pipelines. Notice, however, that this depends on the level of complexity of the method analyzing a specific trustworthy property. For example, the explainability of AI models (through methods like LIME, SHAP, and occlusion sensitivity) is measured by looking at how data inputs influence model outputs, requiring a complete overview of the whole pipeline execution. AI sensors are expected to interact between them, suggesting that by equipping them with further autonomy, it is possible to balance the trust in applications automatically.[10]

Furthermore, once instrumented, the configuration of an AI sensor plays a crucial role in determining the level of trustworthiness in monitoring. The sampling rate directly affects energy consumption and application performance, requiring optimal sampling for improved user experience. While it may seem feasible to sample the AI model every time it updates, the risk of adversarial attacks or induced changes persists over time, requiring frequent model assessment and analysis. Consequently, selecting the optimal sampling frequency for AI sensors remains an ongoing challenge, necessitating further research across various applications. Once sampled, however, the quality of data collected by AI sensors can create several commercialization opportunities. AI sensors yielding data that align well with both legal and technical requirements can gain a competitive edge in the market. This can also create opportunities for certifying AI sensors, facilitating easier auditing and accountability for trustworthy AI software. Certified AI sensors can allow developers to focus more on implementing application-specific functionality rather than evaluating trustworthiness properties.

Dashboard integration and usage

Once sensors are instrumented, measurements can be continuously extracted from AI models, and these can then be presented to users or any stakeholders in dashboards.[10] By using the dashboards, stakeholders can visualize critical aspects that influence the inference behavior of AI models. For example, the level of fairness, robustness, and resilience, to mention a few. Through dashboard inspection, individuals relying on AI models can be aware of the limitations and scope of the decision support provided by AI models. Ultimately, dashboards can support humans in deciding whether or not to use AI to aid with a particular task. As mentioned earlier, effectively presenting trustworthy results is crucial for communicating important AI characteristics to users. The method of presentation, however, depends on the specific type of application being used. Another key challenge that emerges when interacting with AI models through an AI dashboard is the type of device. AI dashboards have to be designed for different types of device characteristics and continuous cross-device interactions—beyond simple screen size. For example, an AI dashboard for a smartwatch may be visualized instead in a smartphone rather than in the smartwatch itself.[15] This is to avoid users misunderstanding information in the dashboard, but it requires designing AI dashboards to fit into multidevice usage patterns. Another example is a self-driving car; a user may pair their personal device with the AI dashboard of the car temporally, such that the user can be aware of the capabilities of the car for navigation.

Human oversight

AI dashboards can also be doors for interacting with AI models. As a part of the EU AI Act, humans play a critical role in overseeing the behavior of AI. However, interacting with AI models is a difficult task, especially when tuning AI models. Human intervention in AI tuning can negatively impact performance by introducing biases or opening back doors based on model recommendations. Thus, a key challenge is to abstract the characteristics and functionality of AI models in a clear and concise form to individuals. This abstraction also has to consider the interaction of AI models with different groups of (stakeholder) users. Here, a group depicts users with different levels of expertise or domain knowledge. This hierarchy also depicts the level of involvement that humans have with the AI tuning. For example, end users may provide feedback, but implementing it requires a different group with specialized skills and domain knowledge. Advancements in large language model (LLM) technologies can aid in this matter, providing an adaptive way to generate explanations for different types of users. Indeed, prompts tailored with domain-specific terminology can be created to communicate with each stakeholder.

Additionally, interaction between AI sensors can also be supported through LLM interfaces, meaning that negotiation happens through natural language interactions. This way, individuals can also have a way of troubleshooting AI behavior just by inspecting dialogue-like conversations. Negotiation between AI-based chatbots has been investigated and demonstrated over the years.[11] Besides this, another key challenge is to determine what changes can be applied to the model by individuals: for instance, removing personal data from the training dataset, changing the machine learning algorithm, hyperparameter tuning of the models (optimizing inference performance), or simply adding/referring new data to the model, among others. This is a critical challenge to overcome as AI models have to support the individual needs of users while preserving general values from groups and society. Otherwise, conflicts on AI usage may arise, halting everyday activities and human processes.

Privacy-preserving and secure monitoring

AI models can be adversely affected by induced and noninduced changes at any stage of their construction pipeline. Noninduced changes emerge from unintentional situations where the data are hampered as they are collected and prepared for storage: for instance, an image corrupted by a camera failure. Similarly, induced changes arise from the intentional manipulation of the data (adversarial attacks). Since analyzing the trustworthiness of AI requires access to the AI model, its dataset, and its pipeline, it is then important to protect them against intentional attacks. Thus, a key challenge is to guarantee that the continuous monitoring of trustworthy properties is conducted in a secure manner.[20] Existing methods based on multiparty computation, homomorphic encryption, and trusted execution environments (TEEs) could be adopted in this matter. Integrating these mechanisms within the architectures, however, requires managing extra computation overhead in the analysis as well as solving several technological limitations to achieve scalable solutions. For instance, while TEEs are currently available to aid in secure computation, they have several limitations regarding the specific characteristics in software runtime execution, for example, programming language, dependencies, and storage to mention the most common.

Legal and technical trustworthiness

Defined regulatory trustworthiness differs when implemented in practice. Indeed, characterizing and measuring trustworthiness in AI is an ongoing process. Several works have developed and proposed different technical methods on how to quantify each aspect of trustworthiness. For instance, several different methods have been proposed to measure the explainability (LIME, SHAP, and Grad-CAM, among others), fairness, and resilience of AI models. Currently, however, there is a clear mismatch between legal/ethical and technical requirements. The EU and U.S. AI Acts have identified requirements to ensure the trustworthiness of AI. Moreover, international initiatives and projects, such as open source SHAPASH, the PwC AI trust index, Microsoft's AI trust and transparency, IBM's AI Fairness 360, and Open AI's AI Impact Assessment, have defined trustworthiness and identified their respective properties. Likewise, EU projects, such as EU TRUST-AI (https://trustai.eu/), EU SPATIAL (https://spatial-h2020.eu/), and EU TAILOR (https://tailor-network.eu/), have also proposed principles and guidelines to ensure trustworthiness in AI development practices.

While there is a clear overlap between all these works, a key challenge that remains unexplored is identifying the essential requirements of trustworthiness. While the assumption is that the EU regulatory approach (properly implemented) could ensure the trustworthiness of AI technologies, it is important that these solutions are interoperable acceptable and manageable options in other legal/economic environments. More importantly, mapping legal/ethical to technical requirements is a critical challenge to identify the limitations and implications of trustworthiness in practice. This can potentially lead to concrete procedures on how AI sensors are constructed and instrumented. Moreover, standard specifications of AI dashboards can also be adopted, such that individuals have a clear understanding of AI even in different geographical and legal/economic environments.

RISKS TO PREDICTION

AI pipelines are a part of larger systems. This suggests that all trustworthy AI properties are not achievable just by examining AI-related components. For instance, security is a property defined by trustworthiness, but securing a large system is a general task carried out for the overall underlying infrastructure and ignores whether AI is present or not in the system. As a result, not all the trustworthy properties can be envisioned only within the scope of AI. In this case, AI sensors can collect measurements to determine the level of security of the entire system, meaning that trustworthy properties are not unique to AI only but they extend to other components of the whole system.

Foundational models are larger models built considering billions of parameters. AI sensors and dashboards embedded into the design stages of these models could easily aid in ensuring that pretrained models are free of biases, secure, and overall trustworthy. Foundational models can, however, pose a big challenge in the use of AI sensors when examining them via post-de facto and verifying their regulatory compliance before using them. Currently, it is unclear to what extent foundational models can be augmented and used within applications without analyzing their retraining and dissecting their inference logic.

While AI dashboards and sensors can provide quantifiable properties about the trustworthiness of AI models, it is difficult to predict whether end users or specific stakeholders would be able to modify/tune the behavior of AI in applications. On the one hand, personalized AI models and the control of an individual's data are key to fostering EU liberties and rights. On the other hand, general models preserving the ethical values and legal/economic requirements of societal groups are key for using AI without conflicts. As a result, AI dashboards can potentially provide insights into effective AI performance, but it is foreseen that changes to tune the behavior of the model would be applicable only by defined authorities. Furthermore, notice also that several technological enablers are currently available to aid in realizing the vision; multiple paths can be followed to build AI sensors and dashboards. However, the use of a specific technology ultimately depends on its rate of development and level of maturity.

Additionally, while it is possible for AI sensors to monitor intentional changes in data, for example, data poisoning, it is unlikely that AI sensors will be used to monitor nonintentional data changes as those are based on situational and management factors. Collecting large volumes of real data that are free of errors and not missing records is unfeasible, and extensive cleaning and preprocessing methods are available to prepare and verify data before training. In parallel to this, generative AI has transformed the use of synthetic data for the training of robust AI models. Generative AI can now be used to augment and enrich scarce datasets, improving the overall decision making of AI models. While the use of generative AI is foreseen to continue and become a standard practice in AI developments, AI sensors and dashboards can foster its safe usage by communicating to users first the quantifiable amount of synthetic data used in the model inference process and second the sources used in the generative creation of the dataset used for training: for instance, text transformed into images or vice versa.

Lastly, it is expected that any application implementing AI functionality is equipped with AI sensors and dashboards. While AI sensors can follow standard guidelines for their instrumentation in software applications, AI dashboards require integration based on the type of application. For instance, AI dashboards in Metaverse applications can be interfaces that are part of the virtual experience, whereas wearable applications require interfaces to be designed for a variety of personal devices. Besides this, it is also possible for users to take for granted the behavior of AI over time. This means that trust in AI is by default expected, and AI dashboards are not frequently checked by individuals. AI dashboards, however, are still required to facilitate the verifying and auditing of AI-based applications before they are released to the public. Moreover, AI dashboards can enable faster response times and proactive decisions when facing cyberattacks.

New regulatory requirements for the development of AI are ensuring the trustworthiness of the technology for its usage in everyday applications. To further strengthen the liberties and rights of individuals when interacting with AI, in this article, we predict a research vision of AI sensors and dashboards. The first gauges and characterizes the behavior of AI models and their evolving trustworthy properties, whereas the latter introduces human-in-the-loop supervision and control to tune and monitor the behavior of AI with human support. We highlighted how modern applications can benefit from AI sensors and dashboards and described the technical research challenges that have to be fulfilled to achieve our vision. C

ABOUT THE AUTHOR

HUBER FLORES is an associate professor at the Institute of Computer Science, University of Tartu, 50090 Tartu, Estonia. His research interests are distributed, mobile, and pervasive computing systems. Flores received a Ph.D. in computer science from the University of Tartu. Contact him at huber.flores@ut.ee.

ACKNOWLEDGMENT

This research is part of a SPATIAL project that has received funding from the European Union's Horizon 2020 research and innovation program under Grant 101021808.

REFERENCES

1. T. Babina et al., "Artificial intelligence, firm growth, and product innovation," *J. Financial Econ.*, vol. 151, Jan. 2024, Art. no. 103745, doi: 10.1016/j.jfineco.2023.103745.
2. S. Herbold et al., "A large-scale comparison of human-written versus chatGPT-generated essays," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 18617, doi: 10.1038/s41598-023-45644-9.
3. A. Goldfarb, "Pause artificial intelligence research? Understanding AI policy challenges," *Canadian J. Econ./Revue Canadienne D'économique*, early access, 2024, doi: 10.1111/caje.12705.
4. B. Li et al., "Trustworthy AI: From principles to practices," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–46, 2023, doi: 10.1145/3555803.
5. "European approach to artificial intelligence." European Commission. Accessed: Mar. 1, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
6. "Executive order (EO) 13960." CIO.gov. Accessed: Mar. 1, 2024. [Online]. Available: https://www.cio.gov/policies-and-priorities/Executive-Order-13960-AI-Use-Case-Inventories-Reference
7. "Interim measures for the management of generative artificial intelligence services." CAC. Accessed: Mar. 1, 2024. [Online]. Available: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm
8. J. M. Wing, "Trustworthy AI," *Commun. ACM*, vol. 64, no. 10, pp. 64–71, 2021, doi: 10.1145/3448248.
9. A. H. Celdran et al., "A framework quantifying trustworthiness of supervised machine and deep learning models," in *Proc. AAI SafeAI2023 Workshop*, 2023, pp. 2938–2948.
10. Y. Wang, "Balancing trustworthiness and efficiency in artificial intelligence systems: An analysis of tradeoffs and strategies," *IEEE Internet Comput.*, vol. 27, no. 6, pp. 8–12, Nov./Dec. 2023, doi: 10.1109/MIC.2023.3303031.
11. S. Chen et al., "An intelligent chatbot for negotiation dialogues," in *Proc. IEEE Smartworld, Ubiquitous Intell. Comput., Scalable Comput. Commun., Digit. Twin, Privacy Comput., Metaverse, Auton. Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta)*, Haikou, China, 2022, pp. 1172–1177, doi: 10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00168.
12. K. Katevas et al., "FLaaS—Enabling practical federated learning on mobile environments," in *Proc. ACM 20th Annu. Int. Conf. Mobile Syst., Appl. Services (MobiSys)*, 2022, pp. 605–606, doi: 10.1145/3498361.3539693.
13. D. Martin, N. Kühl, and G. Satzger, "Virtual sensors," *Bus. Inform. Syst. Eng.*, vol. 63, no. 3, pp. 315–323, 2021, doi: 10.1007/s12599-021-00689-w.
14. C. B. Fernandez et al., "Implementing GDPR for mobile and ubiquitous computing," in *Proc. ACM 23rd Annu. Int. Workshop Mobile Comput. Syst. Appl. (HotMobile)*, 2022, pp. 88–94, doi: 10.1145/3508396.3512880.
15. S. Park et al., "AdaM: Adapting multi-user interfaces for collaborative environments in real-time," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, 2018, pp. 1–14, doi: 10.1145/3173574.3173758.
16. E. Frachtenberg, "Practical drone delivery," *Computer*, vol. 52, no. 12, pp. 53–57, Dec. 2019, doi: 10.1109/MC.2019.2942290.
17. B. S. Miguel et al., "Putting accountability of AI systems into practice," in *Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2021, pp. 5276–5278, doi: 10.24963/ijcai.2020/768.
18. H. Ning et al., "A survey on the metaverse: The state-of-the-art, technologies, applications, and challenges," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14,671–14,688, Aug. 2023, doi: 10.1109/JIOT.2023.3278329.
19. K. Cui et al., "GenCo: Generative co-training for generative adversarial networks with limited data," *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 1, pp. 499–507, 2022, doi: 10.1609/aaai.v36i1.19928.
20. F. Mo et al., "DarkneTZ: Towards model privacy at the edge using trusted execution environments," in *Proc. 18th Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2020, pp. 161–174, doi: 10.1145/3386901.3388946.