Special #4

Bridging Science, Ethics, and Society: The Ethical Management Framework of the PRAESIIDIUM Project

Roberto Larcher, Julia Mader, Gianmarco Veruggio

The **PRAESIIDIUM** project brings together researchers, clinicians, engineers, and ethicists to explore how artificial intelligence can support the early detection and management of prediabetes, while ensuring that innovation remains firmly grounded in ethical responsibility. This article presents a conversation among three key members of the consortium, reflecting on **the ethical challenges and lessons learned throughout the project**. From questions of data ownership, privacy, and trust, to the human impact of predictive models and the evolving relationship between patients, clinicians, and intelligent systems, their dialogue offers a rich perspective on what it means to design AI that truly cares.

Gianmarco Veruggio is a robotic scientist and a leading expert in the ethical and societal implications of science and technology. He is affiliated with the **School of Robotics**, which he co-founded in 2000 together with Fiorella Operto, with the aim

of exploring the complex relationships between society, science, and technology. Formerly the **Director of the Genoa Unit of CNR-IEIIT** (the Institute of Electronics, Computer and Telecommunication Engineering of the Italian National Research Council), he has combined a strong technical background with a deep interest in the human and ethical dimensions of innovation.

Within his current projects, such as **PRAESIIDIUM**; **Veruggio** contributes primarily as an expert on ethics and the societal impact of emerging technologies. His work focuses on analyzing ethical issues in close collaboration with technical and clinical partners, striving to bridge the worlds of science, philosophy, and technology. In this role, he supports project partners in addressing ethical challenges and developing comprehensive ethical self-assessments, ensuring that technological advancement remains aligned with human values and social responsibility.

Julia Mader is a Professor of **Medical Technology at the Medical University of Graz**, where she also serves as **Associate Professor of Medicine in the Division of Endocrinology and Diabetology and as Deputy Head of the Diabetes Outpatient Clinic.**

Trained in medical diagnostics, Professor **Mader's** research and clinical work focus on innovative approaches to diabetes care and metabolic disorders. At the **Medical University of Graz**, she plays a central role in overseeing clinical trials and coordinating the recruitment of participants for structured interviews within collaborative research projects.

Her work combines deep clinical expertise with a strong interest in technological innovation, aiming to bridge the gap between medical research, patient care, and the development of new diagnostic tools that enhance the management and understanding of chronic diseases such as diabetes.

Roberto Larcher is a Research Engineer at **Spindox Labs**, where he develops artificial intelligence solutions across various fields. In the **PRAESIIDIUM** project, he serves as the Technical Project Leader, ensuring the soundness and coherence of all technical components to achieve the project's final objectives. He also leads a Work Package, which focuses on the development of the integrated platform that consolidates the project's technological results into a unified final product.

Larcher's work within **PRAESIIDIUM** requires close collaboration with clinicians, computer scientists, and roboticists to ensure that AI models are both technically robust and trustworthy for clinical use. This involves managing data collected through ICT devices and ensuring that the resulting technologies are reliable, ethical, and aligned with the needs of healthcare professionals and patients alike.

PRAESIIDIUM Ethical Management

The **PRAESIIDIUM** project integrates advanced artificial intelligence and robotic technologies to address a serious medical condition, while also placing strong emphasis on the ethical implications of innovation in healthcare.

Observing, Reflecting, and Engaging

From the outset, the ethical team participated in most project activities as observers and interviewers, providing reflections and proposing ethical considerations as the work evolved. Their methodology was inspired by key European frameworks such as the **ALTAI** (Assessment List for Trustworthy Artificial Intelligence), focusing on principles like human agency, privacy, transparency, and accountability.

This structured approach involved analyzing crucial dimensions such as privacy, by developing a questionnaire to collect the consortium partners' perceptions and attitudes toward emerging ethical challenges. Through this process, the team gathered valuable insights into how trust, data use, and human responsibility intersect in an AI-driven healthcare context.

AI Trustworthiness and Data Ethics

A central theme in **PRAESIIDIUM's** ethical work was AI trustworthiness. As **Gianmarco Veruggio** emphasized, the explainability of artificial intelligence remains one of the most complex and essential challenges in developing ethical AI systems. Understanding how and why an AI model produces its results is key to ensuring clinicians' and patients' confidence.

Closely connected were issues of data privacy and quality. The models relied on datasets from multiple sources — some anonymized, others not — which raised concerns about data protection and anonymization. Although these challenges were carefully managed, they required continuous ethical oversight. Additionally, the quality and nature of the data were ethically relevant, especially regarding gender, ethnicity, and participants' differing levels of understanding. These factors highlighted the need for fairness, inclusiveness, and transparency throughout the project.

A Trustworthy-by-Design Approach

As **Roberto Larcher** explained**, PRAESIIDIUM** adopted a trustworthy-by-design approach, embedding ethical reflection directly into the model development process rather than addressing it afterward. From the very beginning, the team defined a set of desired properties for the AI models, acknowledging that not all could be achieved immediately but identifying critical points and creating roadmaps for improvement.

For example, the heterogeneity of people who could benefit from the system was considered early on. This was reflected in the clinical study through factors such as sex at birth, though it proved more difficult to address ethnicity due to practical and economic limitations. The team emphasized that

awareness of such limitations is essential: developers must recognize when their models may apply only to certain groups and plan to expand inclusivity over time.

Privacy and Data Responsibility

Regarding data privacy, **PRAESIIDIUM** went beyond standard GDPR compliance. During data collection, participants used wearable devices such as Fitbits, systems that operate outside the project's IoT infrastructure and thus beyond direct data management control. To address this, the team ensured that participants were thoroughly informed about how their data would be handled, including potential risks and security measures.

A detailed report on data breaches and device security was prepared, assessing possible vulnerabilities and outlining what would happen in the event of a data breach. The conclusion was clear: when full control over data cannot be guaranteed, transparency and participant awareness become essential. Participants must understand and consciously accept the risks, acknowledging that while some risks exist, the benefits of participation outweigh them.

Explainability and Human-Centered Design

Larcher also highlighted the importance of model explainability as a component of trustworthiness. Many AI models produce results without explaining how those results were obtained. To address this, the **PRAESIIDIUM** team implemented techniques such as counterfactual analysis, which explores "what if" scenarios. For example, when predicting the risk of developing diabetes, the system can also suggest preventive measures, such as a minimum level of physical activity to maintain a healthy state.

This approach not only improves understanding but also mitigates anxiety that could arise from predictive outcomes. By pairing each result with practical guidance, the system empowers users to take informed action, turning predictions into opportunities for health improvement rather than sources of concern.

Responsibility and Human Oversight

Finally, both ethical and technical teams agreed on a fundamental principle: human responsibility must remain central. While AI can support diagnosis and monitoring, the ultimate interpretation and decision-making must always rest with clinicians. This ensures accountability, preserves human judgment, and reinforces the ethical commitment that technology must serve, never replace, the human dimension of healthcare.

Ethics Beyond Committees

Julia Mader noted that, unlike many other European projects where ethical issues are limited to the oversight of ethics committees, **PRAESIIDIUM** took a broader and more participatory approach by engaging the entire consortium step by step in the ethical reflection process.

Having worked in several multidisciplinary projects before, **Julia** explained that such collaboration among clinicians, engineers, and data scientists was not new to her team. However, **PRAESIIDIUM** stood out for the way it extended ethical responsibility beyond formal committees, embedding it into the daily workflow of all partners. One of the main challenges she recalled was that the project had to adapt its study design over time. Originally intended to include both data collection and a clinical trial, **PRAESIIDIUM** ultimately focused only on data collection, a change that required careful justification to the ethics committee. Nevertheless, the process was managed effectively, thanks to the team's familiarity with regulatory procedures.

Lessons from Collaboration

Professor Mader emphasized that early in the project, differences emerged between clinicians and technical partners regarding their understanding of data privacy and ethics. Over time, through dialogue and collaboration, these differences diminished as the partners developed a shared ethical framework. In her view, this collaborative process, fostering mutual understanding between disciplines, is essential for any future transdisciplinary project, and should be built into the planning phase rather than left to evolve during execution.

From a technical standpoint, **Roberto Larcher** reflected that working closely with clinicians was a completely new experience for him. He was surprised by how different it is to collaborate with clinicians compared to engineers, not only in communication styles, but also in expectations, data handling practices, and timelines. Among the main challenges were data transfers and the definition of responsibilities between partners. **Larcher** noted that if these responsibilities and procedures are clearly defined from the very beginning of the project, and included in the consortium agreement, there is no need for additional arrangements later on. Having organized everything up front greatly simplifies the process, avoids misunderstandings, and prevents delays.

Roberto observed that this was one of the key lessons learned during **PRAESIIDIUM**: establishing clear, shared operational and ethical frameworks at the project's inception is essential to ensure smooth collaboration and effective management of sensitive data.

Trust, Data Ownership, and the Perception of Control

Trust in artificial intelligence within healthcare, as both **Julia and Roberto** noted, is not immediate but develops progressively through experience and verifiable results. Julia explained that in diabetes care, clinicians are already accustomed to trusting mathematical models, for instance, those used in automated insulin delivery systems. However, in **PRAESIIDIUM,** no patients were directly exposed to the developed models, making it difficult to form a practical assessment of their reliability. She emphasized that trust must be built step by step, typically through pilot phases under controlled conditions, where safety and efficacy are closely monitored before larger-scale adoption.

Roberto compared this gradual process to the introduction of a new drug, where trust grows with real-world validation. One challenge within **PRAESIIDIUM** was the temporal gap between model forecasts and the duration of clinical studies, while the models aimed to predict health outcomes over several years, the trials spanned only a few months. This underscored the need for a long-term roadmap to evaluate how much these models can be trusted and refined as new data become available.

The use of wearable devices, such as Fitbits and continuous glucose monitors, raised additional questions around privacy and user comfort. Focus group participants, who generally possessed high digital literacy, expressed overall confidence in using such devices and recognized their potential benefits. Nevertheless, some participants, often influenced by relatives or peers, voiced concerns about constant surveillance, feeling uneasy about being monitored around the clock. As **Roberto** highlighted, these reactions underline the importance of clear communication and informed consent, ensuring participants understand the scope of monitoring and the safeguards in place.

He further noted that, while the project's communication efforts successfully alleviated much of this distrust, residual fears persist and should be addressed proactively. One proposal is to limit reliance on devices that store data in third-party cloud services, opting instead for systems managed directly by research institutions. This could help participants feel more confident about who controls their data. **Roberto** also observed that some individuals may have felt there was an imbalance between the information they provided and the benefits they received, a reminder that ethical participation in research requires a perception of fairness and meaningful contribution.

Gianmarco expanded on this by pointing out a broader societal concern: the increasing sense among citizens of being constantly watched and digitally controlled. He observed that as more of our daily interactions occur through screens, smartphones, computers, or voice interfaces, many people feel that their lives are progressively shifting toward a "Matrix-like" reality, where data serve as fuel for corporate or systemic control. This general anxiety about pervasive surveillance could negatively influence people's willingness to participate in digital health monitoring, even when the purpose is benevolent**.** **Gianmarco** argued that such perceptions must be addressed not only ethically, but also socially and politically, to ensure that technological innovation does not deepen public mistrust.

Julia added that one of the major obstacles to trust lies in the behavior of large technology and medical device companies. For example, in the field of glucose monitoring, these companies often act as de facto data owners, even though EU law states that individuals sharing the data should retain ownership. In practice, users frequently face complex and restrictive processes to access their own data, and by agreeing to use a device, they must also consent to the company analyzing it. This contradiction between regulation and reality undermines public confidence. **Julia** noted that some newer companies, such as RO, only provide PDF summaries rather than raw data files, forcing clinicians to negotiate access to the underlying data on behalf of patients. Such practices, she argued, erode trust by making users feel dispossessed of their own information.

Both **Gianmarco and Julia** concluded that true ethical progress in AI-driven healthcare depends not only on technical reliability and data protection compliance but also on structural transparency, user empowerment, and social awareness. Building and maintaining trust will require continuous dialogue

among clinicians, researchers, patients, and institutions, as well as stronger accountability mechanisms for private companies handling sensitive health data.

Knowing the Future: Ethics of Awareness, Care, and Choice

An additional layer of ethical reflection emerging from **PRAESIIDIUM** concerns the question: do people really want to know their future? As **Julia** noted, for most individuals, particularly those who are still healthy, the answer is likely yes, *as long as knowledge empowers them to act.* People who are already attentive to their health and lifestyle are generally open to receiving predictive information, provided that the benefits are tangible and personal, and that data ownership and use are fully transparent.

However, this trust is conditional. **Prof Mader** emphasized that people will only be willing to share their information if they are certain, it will not be used against them, for example, to raise insurance premiums, deny coverage, or influence credit and employment decisions. Predictive data must therefore be handled under strict confidentiality, with no sharing outside authorized research contexts. For AI-assisted healthcare to gain lasting legitimacy, ethical guarantees of privacy, non-discrimination, and data sovereignty must be embedded from the outset.

Larcher agreed, adding that the willingness to know the future depends strongly on whether one can actually do something about it. Predictive insights are empowering when they enable prevention or behavioral change but can be psychologically burdensome when no intervention is possible. For example, learning about a high likelihood of developing Alzheimer's disease, a condition with limited preventive options, might cause distress rather than benefit. This highlights the ethical need for choice and proportionality: people should retain the right *not to know* if they prefer.

At the same time, Roberto pointed out that understanding the real costs of late-stage care could motivate more people to engage in preventive approaches. Many individuals underestimate the economic and social impact of chronic diseases, assuming that future treatments will always be available. Yet the money spent on managing avoidable conditions could, ethically, be redirected to those with greater medical needs. In this sense, **PRAESIIDIUM** has helped participants become more aware of the broader implications of prevention, but further education and dialogue remain essential.

From the **PRAESIIDIUM** focus groups conducted in Graz and Lausanne, it was observed another fascinating dimension: participants expressed a genuine sense of being "cared for" simultaneously by clinicians and by AI. They perceived artificial intelligence not as a distant or cold mechanism, but as a cooperative system, one that *looked after them*, monitored their well-being, and worked alongside doctors. This represents a subtle but profound cultural shift: people are beginning to attribute human-like qualities to AI systems, seeing them as empathetic partners in care.

This anthropomorphic perception, according **to Gianmarco's** point of view, could become an opportunity, if guided ethically. *If AI is seen as caring about patients and their data, it can foster trust and comfort.* But he also cautioned that this dynamic must be balanced carefully to avoid dependence or emotional overattachment. "Let's work to make it something genuinely good," he concluded, "but let's be careful not to let it become overwhelming."

Contattaci:

Contattaci per maggiori informazioni


praesiidium.project@spindox.it


https://praesiidium.spindoxlabs.com


PRAESIIDIUM