HealthTech Blueprint for the Future

Coalition for Innovation, supported by LG NOVA

Jami Diaz, Director Ecosystem Community & Startup Experience
William Barkis, Head of Grand Challenges & Ecosystem Development
Sokwoo Rhee., Executive Vice President, LG Electronics, Head, LG NOVA

Blueprint Chairs

Alex Fang, CleanTech Chair
Sarah Ennis, AI Chair
Alfred Poor, HealthTech Chair

Authors

John Barton, Victor L. Brown, H. Timothy Hsiao, John Hsu, Sylwana Kaźmierska, Ricardo Machado, Ann M. Marcus, Nicholas Matias, Qiana Martin, Alfred Poor, Refael Shamir, Mark Wesson

The views and opinions expressed in the chapters and case studies that follow are those of the authors and do not necessarily reflect the views or positions of any entities they represent.

Senior Editor, Alfred Poor

October 2025

Preamble

The Coalition for Innovation is an initiative hosted by LG NOVA that creates the opportunity for innovators, entrepreneurs, and business leaders across sectors to come together to collaborate on important topics in technology to drive impact. The end goal: together we can leverage our collective knowledge to advance important work that drives positive impact in our communities and the world. The simple vision is that we can be stronger together and increase our individual and collective impact on the world through collaboration.

This "Blueprint for the Future" document (henceforth: "Blueprint") defines a vision for the future through which technology innovation can improve the lives of people, their communities, and the planet. The goal is to lay out a vision and potentially provide the framework to start taking action in the areas of interest for the members of the Coalition. The chapters in this Blueprint are intended to be a "Big Tent" in which many diverse perspectives and interests and different approaches to impact can come together. Hence, the structure of the Blueprint is intended to be as inclusive as possible in which different chapters of the Blueprint focus on different topic areas, written by different authors with individual perspectives that may be less widely supported by the group.

Participation in the Coalition at large and authorship of the overall Blueprint document does not imply endorsement of the ideas of any specific chapter but rather acknowledges a contribution to the discussion and general engagement in the Coalition process that led to the publication of this Blueprint.

All contributors will be listed as "Authors" of the Blueprint in alphabetical order. The Co-Chairs for each Coalition will be listed as "Editors" also in alphabetical order. Authorship will include each individual author's name along with optional title and optional organization at the author's discretion. Each chapter will list only the subset of participants that meaningfully contributed to that chapter. Authorship for chapters will be in rank order based on contribution: the first author(s) will have contributed the most, second author(s) second most, and so on. Equal contributions at each level will be listed as "Co-Authors"; if two or more authors contributed the most and contributed equally, they will be noted with an asterisk as "Co-First Authors". If two authors contributed second-most and equally, they will be listed as "Co-Second Authors" and so on.

The Blueprint document itself, as the work of the group, is licensed under the Creative Commons Attribution 4.0 (aka "BY") International License: https://creativecommons.org/licenses/by/4.0/. Because of our commitment to openness, you are free to share and adapt the Blueprint with attribution (as more fully described in the CC BY 4.0 license).

The Coalition is intended to be a community-driven activity and where possible governance will be by majority vote of each domain group. Specifically, each Coalition will decide which topics are included as chapters by majority vote of the group. The approach is intended to be inclusive so we will ask that topics be included unless they are considered by the majority to be significantly out of scope.

We intend for the document to reach a broad, international audience, including:

- People involved in the three technology domains: CleanTech, AI, and HealthTech
- Researchers from academic and private institutions
- Investors
- Students
- Policy creators at the corporate level and all levels of government

Chapter 15: Observability, Transparency, and Responsible AI Use

Authors: Sylwana Kaźmierska, Ann M. Marcus

Observability, Transparency, and Responsible AI Use

In today's rapidly evolving healthcare landscape, AI systems are becoming integral to improving diagnostics, treatment planning, and patient management. However, with these benefits come responsibilities. Terms such as **observability**, **transparency**, and **responsible AI use** are often tossed around, but what do they really mean in practice? And importantly, how can these principles be implemented by AI practitioners, machine learning engineers, clinicians, and healthcare administrators alike?

While these concepts are interrelated, they serve different roles in managing and deploying AI systems effectively:

- **Observability** is the technical foundation that lets us peer into an AI system's inner workings.
- **Transparency** builds on observability by making those internal processes understandable to everyone, from developers to patients.
- **Responsible AI Use** is the overarching principle ensuring that AI not only performs the intended task well but does so ethically and safely.

Together, they form a comprehensive framework that ensures AI is efficient, trustworthy, and aligned with ethical standards in any application – but especially in medical settings.

Let's examine these three principles in more depth.

Observability: Understanding the Inner Workings

Observability is all about understanding the internal state of an AI system through its monitoring and data analysis. In practice, this involves:

Monitoring and Logging: Capturing operational data such as performance metrics and error logs to track how the system behaves over time. For instance, imagine a diagnostic tool that constantly logs its prediction errors; this data can be invaluable for pinpointing when and why a mistake might have occurred.

So, for example, AI models are built around systems that can shift over time due to a variety of factors. These include evolving consumer profiles in the healthcare sector, generational shifts in workplace behavior, and the emergence of new diseases. So, a model that was initially accurate may degrade over time in unpredictable ways. If this is due to "data drift" – when societal changes make the original datasets inaccurate – it is essential to monitor the model's performance to quickly identify any erroneous trends. This allows for timely regeneration of the model or a reevaluation of its foundational principles.

Metrics and Tracing: Establishing quantitative measures (including latency, throughput, and resource usage) and detailed process traces helps identify performance bottlenecks or anomalies. Think of it as the "black box" in an aircraft; it records everything so that if something goes wrong, engineers can diagnose the cause of the problem quickly.

However, the challenge lies in balancing technical depth with user-friendliness. Too much information might overwhelm non-experts, yet the data must be precise enough to account for user variation and foster trust. Healthcare institutions, therefore, need clear protocols for post-deployment audits, real-time alerts for unusual behavior, and safe rollback procedures for models that aren't performing as expected.

Transparency: Making AI Understandable

Transparency in AI means making the processes, assumptions, decisions, and data handling practices clear and accessible to all stakeholders, including developers, clinicians, insurers, regulators, and patients. This concept is especially crucial in healthcare, where trust and accountability are paramount.

Explainability: At its core, transparency involves offering understandable explanations for AI decision-making. For example, consider a diagnostic tool: clinicians should be able to see which factors influenced a particular decision. This not only helps in verifying the accuracy of the diagnosis but also allows healthcare providers to assess whether the AI's reasoning aligns with clinical expertise.

One such tool for understanding data models is a heatmap. It pinpoints the areas that contributed the most to a model's decision process and outcome(s). Decision trees and regression algorithms implement a set of "feature importance" metrics – that should be transparently established by developers, practitioners, and patients – as well indicate which inputs to the model are the ones that most strongly influenced the decision.

One such example of the limitations in AI decision making is the case of skin tone bias in melanoma detection AI. A study published in the professional journal *Dermis* (April 2025) found that melanoma detection models underperform on darker-skinned patients because training datasets lacked diversity. AI carries the risk of reinforcing existing biases in healthcare, largely stemming from the underlying data rather than the AI algorithms themselves. Because AI models are trained on datasets influenced by human decisions and existing inequities, they may inadvertently perpetuate these biases

Education and Training: To fully leverage transparency, non-technical stakeholders may require training. Integrating AI literacy into medical and nursing school curricula will ensure future healthcare providers understand AI's limitations, inherent biases, and the reasoning behind its decisions–though this is a moving target as AI algorithms are constantly being updated. A well-informed workforce is also less likely to blindly trust AI outputs and more likely to critically assess AI-driven decisions–as they should.

By making AI systems more transparent, we not only build trust but also empower all users — especially those directly responsible for patient care — to make informed decisions about integrating AI into their practices and help patients better understand those decisions and even influence the decision-making process and its assumptions.

Responsible AI Use: Ethics, Accountability, and Compliance

The goal of responsible AI use is to ensure that AI models operate safely, ethically, and effectively. In healthcare, this means that AI models must be validated across diverse populations, clinical settings, and geographic regions to prevent biased or unsafe recommendations. Insufficient testing that relies on a limited population has produced unintended results.

Data Provenance and Quality: AI developers must define a framework that outlines where the data comes from, how its quality is ensured, and what measures are in place to detect bias. Without such a framework, AI models may inadvertently amplify existing disparities in healthcare outcomes. For example, an AI system trained predominantly on data from one demographic might not perform as well for another demographic,

One example, albeit from seven years ago (an eternity in AI development) is this **automatic soap dispenser located that did not recognize black hands**, as shown in this video.

A more recent example is this large-scale study published by the National Institute of Health in April 2025 revealed that LLM-based clinical assistants consistently provided less aggressive diagnostic testing recommendations for low-income patients, despite identical clinical details to high-income counterparts. This bias meant wealthy patients were more likely to receive advanced tests like MRIs or CT scans, reflecting systemic healthcare inequities and raising ethical alarms about fairness in triage delivery based on AI decision support.

Accountability: When AI systems make errors — be it a faulty diagnosis or a biased treatment recommendation — we must have clear accountability frameworks. Whether through internal governance, liability laws, or disclaimers accompanying AI-assisted decisions, assigning responsibility helps avoid ethical and legal dilemmas.

A case in point was reported by Verge in August 2025: Google's Med-Gemini model published a research paper in 2024 introducing a serious hallucination in a section on head CT scans in which it created a part of the brain that didn't exist by conflating two terms –"basal ganglia" and "basilar artery"-- into "basilar ganglia." A blog post also reflected the erroneous term. Nobody at Google caught it, in either that paper or a blog post. The error persisted despite review by dozens of experts until a board-certified neurologist / researcher with expertise in AI flagged the mistake. The blog post was quietly edited with no public acknowledgement, but the paper remained unchanged. Google called the incident "a simple misspelling of 'basal ganglia'," but some medical professionals say it's a dangerous error and an example of the limitations of healthcare AI without real-time monitoring or human-in-the-loop checks.

Regulatory Compliance: The landscape of AI regulation is complex and global. AI systems must comply with laws such as HIPAA, GDPR, and emerging local AI regulations while also adhering to ethical standards. A coordinated, cross-disciplinary regulatory approach is needed to avoid fragmented compliance that could hinder innovation and patient safety.

As an example, General Data Protection Regulation (GDPR) requires that healthcare organizations in the European Union (EU), including the National Health Service (NHS) in the UK, comply with AI-centered companies to incorporate advanced tools for data safety. This has led to the introduction of a federated approach in which machine-learning models are trained on data that remains distributed across multiple locations without the need for the system to "see" the data directly. This approach addresses privacy concerns and data security regulations by keeping sensitive data within its original location while still enabling training for the integrated model.

Embracing Standards as Enablers, Not Barriers

It might seem that all these principles — observability, transparency, and responsible AI use — are cumbersome requirements for engineers and healthcare practitioners to follow. But they are there for protection. Consider again the standards in healthcare such as HIPAA (Health Insurance Portability and Accountability Act), which ensure the confidentiality, integrity, and availability of protected health information and facilitates secure electronic data exchange, safeguarding patient privacy. In the same way, unified standards for AI can simplify development and ensure compatibility and trust across various platforms and institutions.

By establishing common ground rules, AI practitioners can focus more on innovation and less on wheel reinvention. Standards not only streamline the development process but also provide a clear roadmap for integrating AI responsibly into healthcare settings. Ultimately, when every stakeholder — from machine-learning engineers to medical professionals — speaks the same language, the path to safer, more effective AI use becomes much clearer.

Author (In order of contribution)

Sylwana Kaźmierska, Senior Data Scientist, Digica
Sylwana Kaźmierska is a Senior Data Scientist with over 8 years of experience developing machine learning solutions for business. She has worked with companies such as T-Mobile, Dyson, TomTom, AMD, and LEGO. Sylwana is also a TEDx speaker and was recognized as one of the Top 10 Women in AI in Poland as she actively explains AI to a broad audience in her home country.

Ann M. Marcus, Director, Ethical Tech & Communications, WeAccel
Ann M. Marcus is a Sonoma-raised, Portland-based communications strategist and ethical technology analyst focused on smart cities, community resilience, and public-interest innovation. She leads the Marcus Consulting Group and serves as director of ethical technology and communications at WeAccel.io, a public-good venture advancing mobility, communications, and energy solutions for communities. Ann has advised public and private organizations—including Cisco, the City of San Leandro, Nikon, AT&T, and InfoWorld—on trust-based data exchange, digital public infrastructure, resilience strategy, AI and more. Her current projects include a California senior evacuation program, a Portland robotics hub, and digital energy resource initiatives with utilities in Portland and the Bay Area.

For more information about the Coalition for Innovation,
including how you can get involved, please visit coalitionforinnovation.com.

View the Next Chapter