The Hidden RoI of AI Ethics:

The Need for Ethical Decision Making in Healthcare to Balance Innovation & Trust

Michele Loi (Ph.D.), Deepak Bansal, Christian Hugo Hoffmann (Ph.D.)

Implementation Partner:

MQ Learning Academy
Zürich, Switzerland

Research Partner:

University of Zürich
Zürich, Switzerland

Innovation cheque supported by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Application no. 71336.1 INNO-SBM

Table of contents

Executive Summary

Our research reveals distinct patterns in how healthcare organizations approach AI implementation, with significant implications for ethics integration and education. Through 11 in-depth interviews spanning pharmaceutical companies, healthcare providers, and AI solution developers, we identified several key insights that challenge current assumptions about AI ethics in healthcare.

Key insight 1: Implementation Distance and Ethical Salience The adoption of AI technologies follows a distinct pattern correlated with distance from direct patient impact. Applications further from immediate patient care - such as molecular discovery and logistics optimization - show faster adoption rates and less stringent oversight. Conversely, direct patient care applications face more careful scrutiny, revealing how perceived ethical salience, rather than actual ethical impact, often drives implementation decisions.

Key insight 2: Conscious Adoption and Regulatory Gaps Healthcare organizations demonstrate high awareness of AI limitations and bias risks, particularly in regulated clinical areas. However, this consciousness creates an unexpected pattern: while heavily regulated medical applications receive robust ethical oversight, administrative and operational AI systems - often equally impactful on patient outcomes - face less scrutiny despite significant downstream ethical implications. This regulatory gap emerges as a critical challenge for comprehensive ethics integration.

Key Insight 3: Organizational Scale and Strategic Adaptation in AI Ethics Organizations demonstrate distinct approaches to AI ethics implementation based on their size and resources. Some large pharmaceutical companies are beginning to invest in comprehensive ethics frameworks with dedicated teams, while smaller entities and startups must operate within resource constraints, focusing on essential safeguards. This resource disparity creates a systemic challenge where organizational size, rather than ethical impact, often determines the robustness of AI governance. Notably, many organizations strategically position their AI solutions to avoid medical device classification, potentially creating blind spots in ethical oversight. This dynamic highlights how resource availability, rather than ethical necessity, shapes the implementation of AI governance structures across the healthcare sector.

Key Insight 4: Regulatory Framework Dynamics The current regulatory landscape reveals significant gaps in AI oversight, particularly in non-medical device applications and administrative systems. While traditional medical devices face rigorous controls, many AI-powered "assistant" tools operate in regulatory grey areas. This dichotomy creates a complex challenge where innovation speed must be balanced against potential automation bias risks, especially in seemingly low-risk applications that may have cascading effects on patient care.

Key Insight 5: Educational and Implementation Needs The research identifies a dual educational challenge: AI solution providers need guidance on ethical development and validation, while healthcare organizations require capabilities to evaluate and implement AI solutions responsibly. This necessity emerges not just from technical requirements, but from

market pressures and investor demands for ethically-sound implementations. The pattern reveals how external stakeholder expectations increasingly drive internal competency development, creating a feedback loop between market demands and organizational capabilities.

Methodology and Sample Our findings emerge from interviews across the healthcare AI spectrum, including:

- Two large pharmaceutical companies
- Six small companies (including an AI solutions provider)
- A hospital research team
- A venture capital firm

This diverse sample enables analysis of how different organizational contexts shape AI ethics implementation approaches and educational needs.

Research Implications These patterns suggest that effective AI ethics integration in healthcare requires going beyond traditional compliance frameworks to address:

- Overlooked ethical implications in seemingly routine applications
- Varying needs across organizational sizes and types
- Gaps between perceived and actual ethical risks
- Strategic balancing of innovation and safety considerations

This research provides a foundation for developing targeted educational strategies and implementation frameworks that address these complex dynamics in healthcare AI adoption.

1. Key definitions

1. **AI Ethics**: The field of study that focuses on the moral implications of AI development, deployment, and use. It encompasses issues such as fairness, explainability, privacy, trust, and the broader societal impact of AI technologies.
2. **Healthcare:** An encompassing term that includes various sectors and activities related to the provision of health services, products, and technologies. In the context of this report, healthcare includes pharmaceutical companies, hospitals and healthcare providers, health technology companies, biotechnology firms, medical device manufacturers, and healthcare insurance providers. This broad definition allows for an examination of AI adoption and its ethical implications across the entire healthcare value chain.
3. **Clinical application:** In this text, "clinical application" refers to the use of AI technologies in direct patient care settings, such as genomic analysis for personalized therapy decisions, resource optimization and length of stay prediction in rehabilitation, and surgical planning and 3D anatomical measurements in cardiac care.
4. **Ethical Decision Making**: The process of evaluating and choosing among alternatives in a manner consistent with ethical principles. In the context of AI, ethical decision making involves considering the potential impacts of AI systems on individuals, groups, and society as a whole, and making choices that prioritize fairness, transparency, accountability, and the well-being of those affected by the

technology. In healthcare AI, we understand it as the systematic process of evaluating AI implementation choices based on their potential impact on patient care, organizational accountability, and societal outcomes, guided by ethical considerations.

5. **AI Maturity**: The extent to which an organization has adopted and integrated AI technologies into its operations, processes, and decision-making. AI maturity can be assessed across various dimensions, such as data availability, technical infrastructure, talent, governance, and strategic alignment.
6. **Regulatory Gaps**: Areas where existing laws, regulations, and policies do not adequately address the unique challenges and risks posed by AI technologies. These gaps can create uncertainty for organizations adopting AI and may require the development of new regulatory frameworks.
7. **AI Governance**: The processes, policies, and structures that organizations put in place to oversee and manage the development, deployment, and use of AI systems. Effective AI governance helps ensure that AI is used in an ethical, transparent, and accountable manner, while also supporting business objectives.

2. Methodology: From Literature Review to Field Research

Our research methodology consisted of two primary stages:

1. Literature review: Identified four key areas of ethical concern in AI - fairness, explainability, privacy, and trust.
2. Expert interview (revealed two critical blind spots in academic discussions:
    - Overemphasis on heavily regulated AI applications; underestimation of organizational AI significance.
    - Insufficient attention to routine administrative AI applications with significant downstream ethical implications).

Our hypotheses and focus were revised based on insights from the expert interview abbreviated here as Int1EthicsExp. Following this insight, we conducted 10 semi-structured interviews across a diverse sample of healthcare organizations. Our overall sample consisted of:

Interview ID	**Organization Type**	**Field**
Int1EthicsExp	Academia and ethics consulting	Ethics Expert
Int2bus	Large Pharmaceutical	Pharmaceutical

Int3RehabResearch	Hospital Research Team	Diagnostics
Int4PharmaSmallExec	Small Company (Swiss)	Pharmaceutical
Int5PharmaBiotechAI	Small Company (Swiss)	Pharmaceutical
Int6StartupRandD	Small Company (Swiss)	Pharmaceutical
Int7VCBiotech	Venture Capital	Biotech Investment
Int8AIProvider	Small Company (Swiss) - AI Solutions	Pharmaceutical
Int9PharmaBranded	Small Company (Swiss)	Pharmaceutical
Int10PharmaBioethics	Large Pharmaceutical	Pharmaceutical
Int11DiagnosticsCEO	Small Company (UK)	Diagnostics

Our interview guide was structured to allow relevant content to emerge naturally, without explicitly prompting a specific focus for ethical considerations. For pharmaceutical companies, we explored AI applications across the value chain, future potential, adoption challenges, and ethics awareness needs. For healthcare technology and medical diagnostics organizations, we investigated current AI usage, potential applications, regulatory requirements, and ethical concerns. Responses were analyzed within a framework examining the application value chain positioning and maturity of each application, current usage patterns, AI potential, adoption obstacles, and opportunities for ethical support.

Validation

Our research validated the following key points:

a) There is a strong case for AI adoption in the Healthcare industry across the value chain

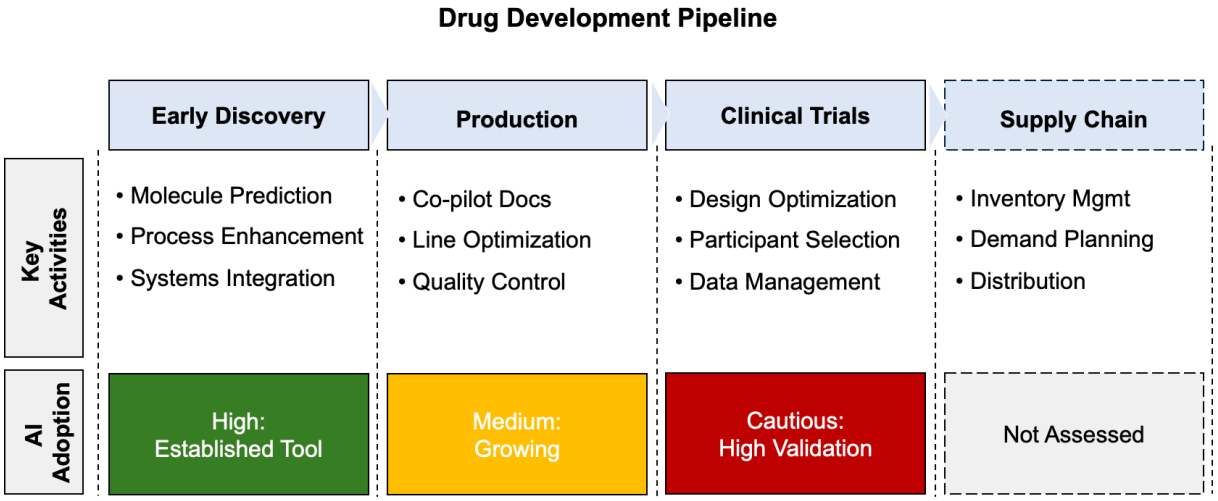
b) AI ethics could be seen as an enabler of prudent AI adoption

c) Education and awareness would play a role in the proactive mitigation of risks while deploying AI technologies

3. Landscape of AI application maturity in the pharma and healthcare industry

3.1 Drug Development Pipeline

AI applications advance drug development across stages, with varying adoption rates based on patient impact. In early discovery, molecule prediction and process enhancement tools are being readily implemented [Int5PharmaBiotechAI]. Production applications focus on co-pilot documentation and line optimization, with expert review and quality control [Int4PharmaSmallExec]. AI adoption in clinical trials is more measured due to ethical considerations and potential risks, focusing on optimized design, participant selection, and document management platforms [Int10PharmaBioethics]. (While supply chain applications likely present significant opportunities for AI implementation, our research did not include expert perspectives from this area of pharmaceutical operations.) Overall, AI implementation is faster in early discovery and process optimization where patient impact is lower, while adoption is more cautious in patient-facing stages.



3.2 Clinical Applications

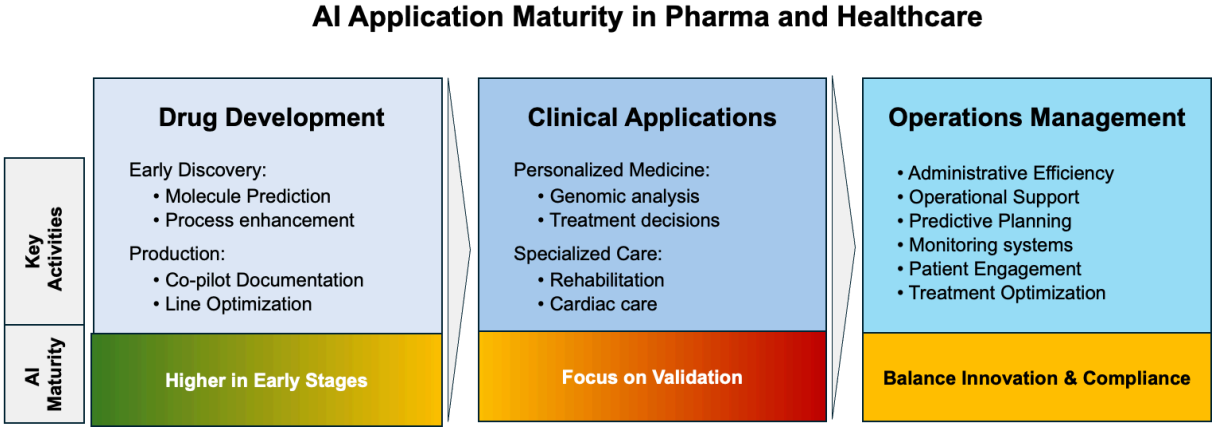
In clinical settings, AI is being explored for personalized therapy decisions, resource optimization, and improving patient care. Genomic analysis tools are demonstrating potential to support individualized treatment while maintaining parameter transparency and quality assurance [Int11DiagnosticsCEO]. In rehabilitation, AI is being used to optimize resources and predict length of stay, benefiting post-surgical care [Int3RehabResearch]. Cardiac care applications reviewed in our study focus on enhancing surgical planning and 3D anatomical measurements, with an emphasis on clinical workflow integration and doctor-centric design [Int6StartupRandD]. Across clinical domains, patient-facing AI applications undergo rigorous validation and slower implementation cycles compared to administrative and resource optimization solutions.

3.3 Operation management

Across the industry, operational AI spans from basic to advanced applications:

- Administrative efficiency: vocal transcription, record summarization, encoding [Int8AIProvider]
- Operational support: AI-assisted workflows, decision recording, data management [Int9PharmaBranded]
- Predictive planning: medical device deployment [Int6StartupRandD]
- Monitoring: vitals alerts, glucose, mental health [Int7VCBiotech]
- Patient engagement: medication management, follow-up, symptom tracking [Int8AIProvider]
- Treatment optimization and selection [Int9PharmaBranded]

While medical device AI has clear regulations, administrative and operational AI often falls into grey areas. Healthcare AI companies frequently respond strategically to this regulatory landscape by positioning their solutions as administrative tools or documentation aids rather than medical devices, targeting workforce gaps and operational inefficiencies. While this approach enables faster innovation, it also creates potential oversight gaps that carry ethical risks. Organizations must balance rapid innovation with responsible practices, particularly in areas where formal oversight frameworks are still emerging.



3.4 Adoption challenges

Healthcare organizations take varied approaches to AI based on their size, capabilities, and regulatory context. Early-stage AI companies prioritize survival and basic compliance until well-funded, while large pharma enterprises establish comprehensive AI governance frameworks [Int4PharmaSmallExec, Int10PharmaBioethics]. Healthcare providers have to balance AI with patient care priorities, considering implementation challenges, demographics, trust, and validation needs [Int3RehabResearch, Int6StartupRandD].

4. Ethical challenges

4.1 Ethical Challenges Across AI Maturity Stages

Healthcare organizations encounter distinct ethical challenges as they progress through different stages of AI adoption. Here are the key considerations at each stage:

Early-stage healthcare organizations:

- Need lightweight ethical guidelines to govern initial AI development
- Must establish data collection and privacy best practices from the outset
- Require strategies to mitigate bias despite working with limited datasets

Mid-stage healthcare organizations:

- Face challenges integrating AI with legacy systems as they expand capabilities
- Experience increased regulatory scrutiny when deploying initial solutions
- Need comprehensive ethical review processes for expanding AI applications
- Require technical guidance on infrastructure integration and data management
- Must navigate complex regulatory landscape and clinical validation requirements
- Need staff training programs focused on human oversight of AI systems

Late-stage healthcare organizations:

- Must balance rapid innovation with patient safety and regulatory compliance
- Need to address potential disparities in how effectively solutions work across different populations
- Require comprehensive ethical frameworks and governance structures
- Must implement robust processes for ongoing monitoring and refinement of deployed AI
- Need strategies to ensure equitable access and effectiveness of AI solutions
- Should collaborate actively with regulators to develop standards and best practices

4.2 Core Ethical Themes: Fairness, Bias, and Trust

Fairness and Bias

The healthcare sector's approach to AI fairness and bias management reveals a sophisticated understanding that aligns with the "Distance from Ethically Salient Impact" principle. In molecular discovery and chemical applications, where AI operates at a greater distance from direct patient impact, bias concerns focus primarily on historical target selection patterns rather than demographic representation [Int5PharmaBiotechAI, Int7VCBiotech]. For instance, pharmaceutical research faces systematic biases in disease target selection, with historical data and model success rates skewed toward conditions prevalent in Western populations and commercially viable markets [Int5PharmaBiotechAI]. This biases cannot be tackled by companies with specific mandates to generate functional molecules for the given targets, as the generation of the bias lies upstream.

On the contrary, clinical applications face immediate challenges with demographic representation, including age-specific limitations and ethnic homogeneity in training data [Int3RehabResearch]. This is particularly evident in rehabilitation contexts, where patient populations may be regionally specific (e.g., Lombardy) or demographically skewed, such as gender distributions reflecting disease prevalence patterns [Int3RehabResearch].

Practical fairness challenges emerge in unexpected ways across different applications. In scheduling systems, for example, the complexity extends beyond traditional demographic considerations to include cultural and religious sensitivities [Int9PharmaBranded]. In medical imaging and procedural planning, such as cardiac valve operations, bias manifests in the technical challenge of standardizing measurements across diverse anatomical presentations [Int6StartupRandD].

A particularly nuanced challenge emerges in applications building upon existing AI systems, especially large language models, where inherited biases must be carefully considered [Int8AIProvider]. This highlights the importance of continuous monitoring and evaluation rather than assuming initial test data adequacy [Int2bus, Int8AIProvider]. The failure of previous high-profile healthcare AI initiatives underscores the critical need for human calibration and contextual understanding in bias management [Int2bus].

These varied experiences across the healthcare AI landscape highlight why fairness and bias management requires specific knowledge in:

- Context-specific bias evaluation methods
- Continuous monitoring protocols
- Cultural and demographic sensitivity
- Integration of human oversight and calibration
- Understanding of inherited system biases

Trust, explanations and monitoring

Established technical approaches to bias management (such as demographic requirements for clinical trials) form a crucial foundation for building stakeholder trust. While mastering the

technical aspects is essential, our research reveals that trust building requires going beyond purely technical solutions. Rather than focusing primarily on making AI systems interpretable, stakeholders across healthcare contexts emphasize more holistic and contextual approaches to trust-building.

Our interviews reveal that healthcare providers prioritize comprehensive quality assurance over potentially controversial XAI methods for achieving transparency in AI systems [Int8AIProvider]. This preference reflects a pragmatic understanding that reliable performance, rather than algorithmic interpretability, often serves as the foundation for trust in healthcare settings. Clinical validation studies and comparative measurement analyses [Int6StartupRandD] serve not just regulatory requirements but establish concrete evidence of reliable performance. This approach acknowledges that understanding how an AI system works may be less crucial than verifying that it works consistently and reliably within defined parameters.

A crucial insight emerged from our interviews regarding the relationship between trust and system limitations. One healthcare provider notably characterized AI reasoning as too "linear" to be fully trusted in complex healthcare contexts [Int9PharmaBranded]. This observation points to a sophisticated understanding that trust in AI systems should be calibrated rather than maximized. This approach suggests that appropriate trust calibration - understanding when and how much to trust AI systems - often proves more valuable than increasing trust in absolute terms.

This understanding of AI limitations serves a dual purpose in trust-building:

1. It helps calibrate expectations and trust levels appropriately across different use contexts
2. It enables more effective integration of human expertise to complement AI constraints
3. It facilitates the development of more robust validation protocols that account for known limitations

Summing up, trust-building has significant implications for both AI implementation and ethics education in healthcare settings. Rather than focusing primarily on making AI systems more interpretable, organizations should consider:

1. Developing comprehensive quality assurance frameworks that demonstrate reliable performance
2. Creating clear communication strategies about system limitations and appropriate use contexts
3. Implementing context-specific validation protocols that address the particular needs of different user groups
4. Establishing ongoing monitoring systems that maintain trust through consistent performance verification
5. Integrating ethical decision-making frameworks that align with specific stakeholder needs and contexts

Organizational Capabilities and Gaps

The ability to navigate context-specific manifestations of bias, develop appropriate trust calibration strategies, and integrate human oversight depends on robust ethical decision-making processes. However, the varied approaches to AI adoption based on organizations' size, capabilities, and regulatory context, as discussed in Section 3.4, also impact their ability to address ethical challenges effectively [Int7VCBiotech, Int10PharmaBioethics, Int11DiagnosticsCEO].

Smaller organizations face particular challenges in ethical AI implementation. Resource constraints and immediate operational pressures can limit their ability to develop sophisticated ethical frameworks beyond basic compliance requirements [Int7VCBiotech, Int11DiagnosticsCEO]. This creates an uneven landscape where ethical oversight capabilities correlate strongly with organizational resources.

Addressing these disparities requires developing flexible approaches to ethical implementation that acknowledge diverse organizational contexts. Such frameworks must be adaptable enough to serve both resource-rich enterprises and smaller players while maintaining consistent ethical standards. This suggests the need for scalable solutions that can grow alongside an organization's capabilities and resources.

5. Key enhancements of ethical decision-making.

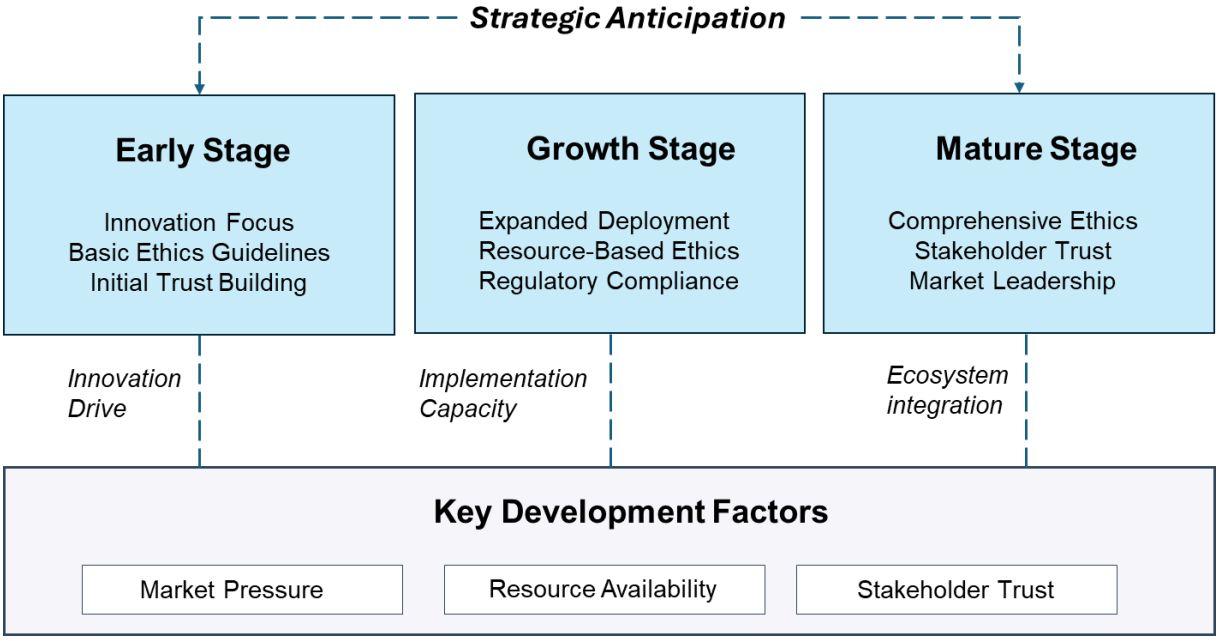
Navigating the complex ethical challenges of healthcare AI requires organizations to cultivate key capabilities. Evidence from expert interviews highlights several vital capability enhancement areas:

1. **Integrating Technical and Ethical Expertise** Simultaneously addressing the technical and ethical aspects of AI implementation necessitates cross-disciplinary collaboration [Int5PharmaBiotechAI].
2. **Strengthening Fairness and Bias Assessment** Building trust in healthcare AI depends on rigorously evaluating systems for fairness and bias alongside conventional performance metrics [Int11DiagnosticsCEO]. Essential capabilities include constructing diverse, representative datasets for AI testing [Int3RehabResearch, Int6StartupRandD]
3. **Adapting Trust-Building Approaches to Context** Calibrating appropriate trust levels in AI requires understanding the nuances of specific healthcare use cases, impact, and risk profiles [Int7VCBiotech]. Organizations must cultivate:
    - Strategies for building trust through transparently communicating AI capabilities and limitations [Int6StartupRandD]
    - Protocols for ensuring appropriate human oversight in AI-augmented decision-making [Int3RehabResearch]
4. **Leveraging External Collaboration** Engaging external capabilities is particularly important for smaller healthcare organizations with limited resources [Int8AIProvider]. Strategies could include:
    - Participating in industry partnerships and collaborative initiatives to share knowledge and resources

    - Contributing to the development of industry standards and best practices
5. **Emphasizing Human-Centered Implementation** Ensuring meaningful human involvement and control is critical for managing ethical risks, particularly in clinical decision support where automation bias is a concern [Int9PharmaBranded].

As healthcare organizations progress through different growth stages, their approach to ethical AI capabilities evolves. As startups and early-stage companies mature and their AI systems become more widely deployed, ethical considerations become increasingly critical for maintaining trust and regulatory compliance [Int5PharmaBiotechAI].

A proactive approach emerges when external stakeholders actively shape ethical capacity development from the earliest stages. Venture capital firms increasingly require robust ethics frameworks as part of their due diligence [Int7VCBiotech], while healthcare providers demand evidence of ethical considerations before pilot deployments [Int3RehabResearch]. This external pressure is complemented by growing support ecosystems - industry consortia providing ethical guidelines, academic partners offering validation frameworks, and specialized consultancies supporting implementation [Int1EthicsExp, Int2bus]. Such external mechanisms may help organizations build ethical capabilities before scaling, when changes are less costly and more effective, although the landscape is hard to navigate. The key is moving from reactive compliance to proactive capability building, where external requirements and support create a positive cycle of ethical development that aligns with business growth [Int10PharmaBioethics]."



Findings summary

The interviews reveal how ethical awareness must extend beyond regulatory compliance, particularly in overlooked administrative applications where AI's ethical implications remain significant. Companies' e\\volution from basic compliance to sophisticated frameworks demonstrates this essential progression, which may reveal gaps and either

under-investments and over-investments at specific key stages in business development. Further research would be needed to understand how the process of ethical know-how acquisition and development can be optimized, potentially through the strategic implementation of AI ethics training, resources, and expert guidance tailored to each stage and organizational context.

Which expert groups within organizations would need Digital or AI ethics education?

The insights from interviews highlight the diverse educational needs across different expert groups within healthcare organizations. Patients, clinicians, and researchers each require tailored AI ethics education to address their specific concerns and roles in the AI deployment process.

For AI solution providers, education should focus on training in design principles for transparent and explainable AI systems, coupled with access to resources such as frameworks for assessing algorithmic bias and expert guidance on regulatory compliance standards. Healthcare organizations must equip clinicians with skills to incorporate AI insights into clinical decisions, emphasizing nuanced judgment, supported by training on the nuances of AI-assisted clinical judgement, resources like case studies illustrating ethical trade-offs, and expert guidance in developing appropriate governance structures for their organization's size and complexity.

What are the major challenges at the expert level within companies that hinder the integration of digital ethics practices?

Our research reveals that experts face multiple interconnected challenges in integrating digital ethics practices. While technical barriers like data quality and system integration pose immediate difficulties, the more subtle challenge lies in calibrating trust appropriately across different AI applications. Experts struggle particularly with applications distant from direct patient impact, where ethical implications are often overlooked despite significant downstream effects. Resource constraints compound these challenges, with smaller organizations lacking dedicated ethics infrastructure and larger ones grappling with complex governance requirements. The absence of clear frameworks for non-medical device applications creates uncertainty in ethical decision-making, while the rapid pace of AI adoption often outstrips the development of ethical guidelines. Additionally, experts face the challenge of balancing innovation demands with ethical considerations, particularly in areas where regulatory guidance is limited or absent. Addressing this requires training to understand these indirect ethical impacts, resources that provide guidelines for identifying collective effects, and expert guidance on ethics frameworks suitable for less-regulated applications.

What specific knowledge and insights in digital and AI ethics would empower experts to make informed and timely decisions?

To empower experts in making informed decisions, AI ethics education should cover key areas such as identifying and mitigating bias, ensuring algorithmic transparency, and navigating regulatory complexities. Case studies illustrating the ethical trade-offs and unintended consequences of AI deployment can help build critical thinking skills.

Entrepreneurship patterns in AI healthcare, such as focusing on patient community tools, scheduling optimization, and documentation processing, highlight the need for strategic education on regulatory compliance and responsible innovation.

For example, a rehabilitation research team developing an AI gait analysis (using artificial intelligence to analyze how a person walks or runs by processing data from sensors, cameras, or wearable devices) would benefit from foundational training in AI ethics principles and privacy considerations, access to resources like bias mitigation tools, and expert guidance through regulatory compliance consultation. Proactively addressing these ethical considerations through targeted education can help build trust with stakeholders and ensure responsible AI deployment.

In conclusion, the interview data reveals a complex interplay between organizational maturity, resource availability, and ethics implementation in healthcare AI. Key patterns emerge around the evolution of ethical frameworks - from basic compliance in early-stage companies to sophisticated governance structures in larger organizations. This progression highlights critical gaps in both regulatory coverage and practical implementation capabilities. Particularly noteworthy is the "regulatory shadow" effect, where highly regulated clinical applications receive robust oversight while less individually impactful administrative AI systems face limited scrutiny, even though they may have significant collective effects, e.g., generating self-reinforcing inequities in healthcare resource distributions. This finding connects directly to our research question about identifying and addressing organizational barriers to ethics integration. The evidence suggests that effective ethics education must extend beyond compliance frameworks to address these overlooked areas of ethical significance. Strategically integrating AI ethics training, resources, and expert guidance throughout the tool development process and across expert levels can help healthcare organizations navigate the complex challenges of AI deployment while maximizing benefits and building stakeholder trust. These insights lay the groundwork for examining specific educational needs across different expert groups and organizational contexts in the following section.

Key message: effective AI ethics education in healthcare requires:

- Tailoring initiatives to the needs of diverse expert and non-expert groups (patients, clinicians, researchers)
- Adapting to varying organizational contexts and resources (start-ups vs. market leaders)
- Addressing key challenges such as distance from ethical impact, conscious adoption barriers, and regulatory gaps
- Empowering experts with knowledge on bias mitigation, transparency, regulatory compliance, and responsible innovation

5. Conclusions:Toward Comprehensive AI Ethics Integration in Healthcare

Our research reveals a complex ethical landscape in healthcare AI, shaped by the interplay of technical factors, organizational contexts, and regulatory frameworks. The findings illuminate how these forces influence ethical decision-making and awareness, exposing critical gaps and opportunities for more effective ethical implementation.

A key insight emerges around the relationship between an AI application's distance from direct patient impact and the speed of its adoption. Applications like molecular discovery and back-office automation, which are further removed from clinical decision-making, face fewer ethical hurdles and thus experience faster uptake. Conversely, AI tools for patient care, clinical decision support, and monitoring systems encounter more scrutiny due to their proximity to patient outcomes, slowing their adoption. This pattern highlights the need for nuanced ethical decision-making frameworks that consider the spectrum of AI applications and their diverse implications.

Organizational size and maturity also play a significant role in shaping ethical practices. Larger, well-resourced organizations can develop comprehensive, multi-year strategies with dedicated ethics teams and governance structures. Startups and early-stage companies, however, often prioritize survival and core operations, leaving ethical decision-making processes minimally developed. Reaching key funding thresholds can catalyze the formalization of ethical practices, including professionalized management and dedicated compliance roles. This evolution underscores the importance of flexible, scalable approaches to ethical decision-making that can adapt to organizational growth.

Critically, the research exposes significant regulatory gaps, particularly in AI applications not classified as medical devices. While healthcare organizations demonstrate high awareness of AI risks in regulated clinical areas, this consciousness doesn't consistently extend to all impactful applications. The findings directly address the grant's key questions:

Expert groups needing AI ethics education: The research reveals a need for ethics education across diverse roles, from AI solution providers to healthcare organization leaders. Deployers' ethical awareness, investors' requirements, and market pressure for ethical implementation emerge as key drivers.

Challenges hindering digital ethics integration: Interviews highlight challenges such as limited ethical decision-making resources in startups, regulatory compliance complexity, and balancing innovation with safety, and the difficulty of addressing "unofficial automation" in administrative AI systems.

Empowering knowledge for informed decisions: The findings suggest a dual focus on educating AI providers and healthcare organizations. Key areas include understanding AI limitations, recognizing bias risks, and navigating regulatory frameworks.

Research Gaps and Future Assessments: The evolution of ethical decision-making capabilities in healthcare AI organizations demonstrates clear patterns of development

across organizational lifecycles, progressing from basic guidelines through resource-based ethics to comprehensive frameworks. However, a significant research gap exists: while we understand the general progression pattern, we lack systematic evidence about how specific funding stages trigger changes in ethical decision-making capabilities and shape organizations' ability to align with social values and exercise ethical foresight. Understanding these funding-ethics dynamics could provide crucial insights for building robust ethical capabilities before scale-up pressures make changes more costly, ultimately determining a startup's capacity to integrate successfully into broader healthcare ecosystems.

In conclusion, our research paints a nuanced picture of the ethical landscape in healthcare AI, revealing the complex interplay of technical, organizational, and regulatory factors. By exposing critical gaps and identifying key educational needs, these findings provide a roadmap for more effective ethical decision-making. As the healthcare sector continues to evolve, a commitment to proactive, context-sensitive ethical frameworks will be essential for realizing AI's transformative potential while safeguarding patient well-being and public trust.

Bibliography

AI Principles. (2017). Future of Life Institute. https://futureoflife.org/ai-principles/

Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? European Journal of Human Genetics, 26(7), 931–936. https://doi.org/10.1038/s41431-018-0160-0

Dave, P. (2024). How AI Can Revolutionize the Pharmaceutical Industry. Journal of Drug Delivery and Therapeutics. https://www.semanticscholar.org/paper/How-AI-Can-Revolutionize-the-Pharmaceutical-Dave/6190b581c463d23982706577433829066ca02536

Department of Health & Social Care - UK Government. (2018, September 5). Initial code of conduct for data-driven health and care technology. GOV.UK. https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

ELSI Advisory Group. (2017, June 12). Ethical Framework for Responsible Data Processing in the Swiss Personalized Health Network. Swiss Personalized Health Network.

Ferretti, A., Schneider, M., & Blasimme, A. (2018). Machine Learning in Medicine: European Data Protection Law Review, 4(3), 320–332. https://doi.org/10.21552/edpl/2018/3/10

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Hagendorff, T. (2020). The Ethics of AI Ethics—An Evaluation of Guidelines. Minds and Machines. https://doi.org/10.1007/s11023-020-09517-8

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kim, J., & Min, M. (2024). From RAG to QA-RAG: Integrating Generative AI for Pharmaceutical Regulatory Compliance Process (No. arXiv:2402.01717). arXiv. https://doi.org/10.48550/arXiv.2402.01717

Loi, M., Heitz, C., & Christen, M. (2020). A Comparative Assessment and Synthesis of Twenty Ethics Codes on AI and Big Data. 2020 7th Swiss Conference on Data Science (SDS), 41–46.

Tummala, S. R., & Gorrepati, N. (2024). AI-driven Predictive Analytics for Drug Stability Studies: Review Article. Journal of Pharma Insights and Research, 2(2), Articolo 2.

Ullagaddi, P. (2024). Leveraging Digital Transformation for Enhanced Risk Mitigation and Compliance in Pharma Manufacturing. Journal of Advances in Medical and Pharmaceutical Sciences, 26(6), Articolo 6. https://doi.org/10.9734/jamps/2024/v26i6697

Université de Montreal. (2017). Montreal Delcaration for a Responsible development of AI. https://www.montrealdeclaration-responsibleai.com/the-declaration

Yadav, S., Singh, A., Singhal, R., & Yadav, J. P. (2024). Revolutionizing drug discovery: The impact of artificial intelligence on advancements in pharmacology and the pharmaceutical industry. Intelligent Pharmacy, 2(3), 367–380. https://doi.org/10.1016/j.ipha.2024.02.009