

THE RESEARCH UNIVERSITY PRESS

International
Computational Law Journal

VOLUME 2 / ISSUE 2 / CURATED SOURCE EDITION

Model for Professionals to Achieve Effective Supervision of Artificial Intelligence

Olga Mack

MIT Computational Law Report · Originally published 2023-08-31

Independent reader edition under CC BY 4.0. This attribution cover was added; original source PDF pages follow unchanged. Original authorship and source identity are retained. MIT, Stanford University, Stanford Law School and CodeX have not sponsored or endorsed this republication.

Source, license and version evidence accompany the online article.

MIT Computational Law Report • Media

Model for Professionals to Achieve Effective Supervision of Artificial Intelligence

Olga Mack

Published on: Aug 31, 2023

URL: https://law.mit.edu/pub/model-for-professionals-to-achieve-effective-supervision-of-artificial-intelligence

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

Artificial Intelligence (AI) has painted a new horizon, breathing life into innovative practices and revolutionizing how professionals navigate tasks. As AI ascends to its prominent role, it's like taming a wild horse – full of promise, but also requiring guidance and oversight. Thoughtful, skillful, well-intended operators are a must. In this article, we demystify the concept of AI supervision and equip professionals with the knowledge to harness the full potential of AI effectively and ethically.

Obligation to Supervise in Different Fields

Supervision is fundamental in many professional fields. It typically falls on individuals in positions of authority who guide, mentor, and ensure the proper execution of tasks. The primary aim of supervision is to ensure adherence to standards, proper guidance, and the professional growth of those supervised.

In professions such as law, medicine, education, finance, construction, law enforcement, and the military, supervision takes on unique forms, but the basic principles are remarkably similar. For instance, supervising attorneys oversee the work of junior staff, ensuring adherence to professional ethics and promoting professional development. Experienced physicians supervise medical residents and interns, overseeing patient care and ensuring compliance with medical standards. Administrators supervise teachers and staff in education, ensuring adherence to curriculum guidelines and maintaining an environment conducive to learning.

The common thread across professions is the welfare of the client and the public. The model rules or regulations in each field aim to ensure this welfare, underscoring the importance of supervision in professional conduct and services delivery.

Frameworks of Supervision: Guiding Professional Development across Diverse Fields

Supervision is a multifaceted concept with roots in diverse fields such as law, education, management, social work, nursing, and psychology. In business, supervisory roles often entail functions like planning, organizing, staffing, directing, and controlling. Similar frameworks are seen in social work, nursing, and psychology, where supervision cycles often include stages such as relationship and skills development, ongoing evaluation, and competency acquisition.

These supervisory frameworks generally guide and develop the skills of budding professionals, e.g., novice teachers under the mentorship of experienced educators, junior managers guided by senior management, or trainee psychologists supervised by licensed psychologists.

Interestingly, supervisory frameworks are finding new applications in our digital world. The advent and integration of AI tools in professional settings give rise to a unique adaptation of traditional supervisory concepts. In this discussion, we present the Four Pillars of Oversight (Oversight, Improvement, Empowerment,

and Harmonization) and the Four Steps of Supervisory Flow (Immersion, Introspection, Evaluation, and Execution) as a framework for professionals to effectively leverage AI tools in their practices. Although we illustrate this framework primarily with lawyers and legal professionals, we can easily apply a similar approach to using AI in education, management, social work, nursing, psychology, and many other fields.

The supervisory duty is a crucial safeguard for both the public and the specific audiences that each profession serves. This protective role manifests in three ways.

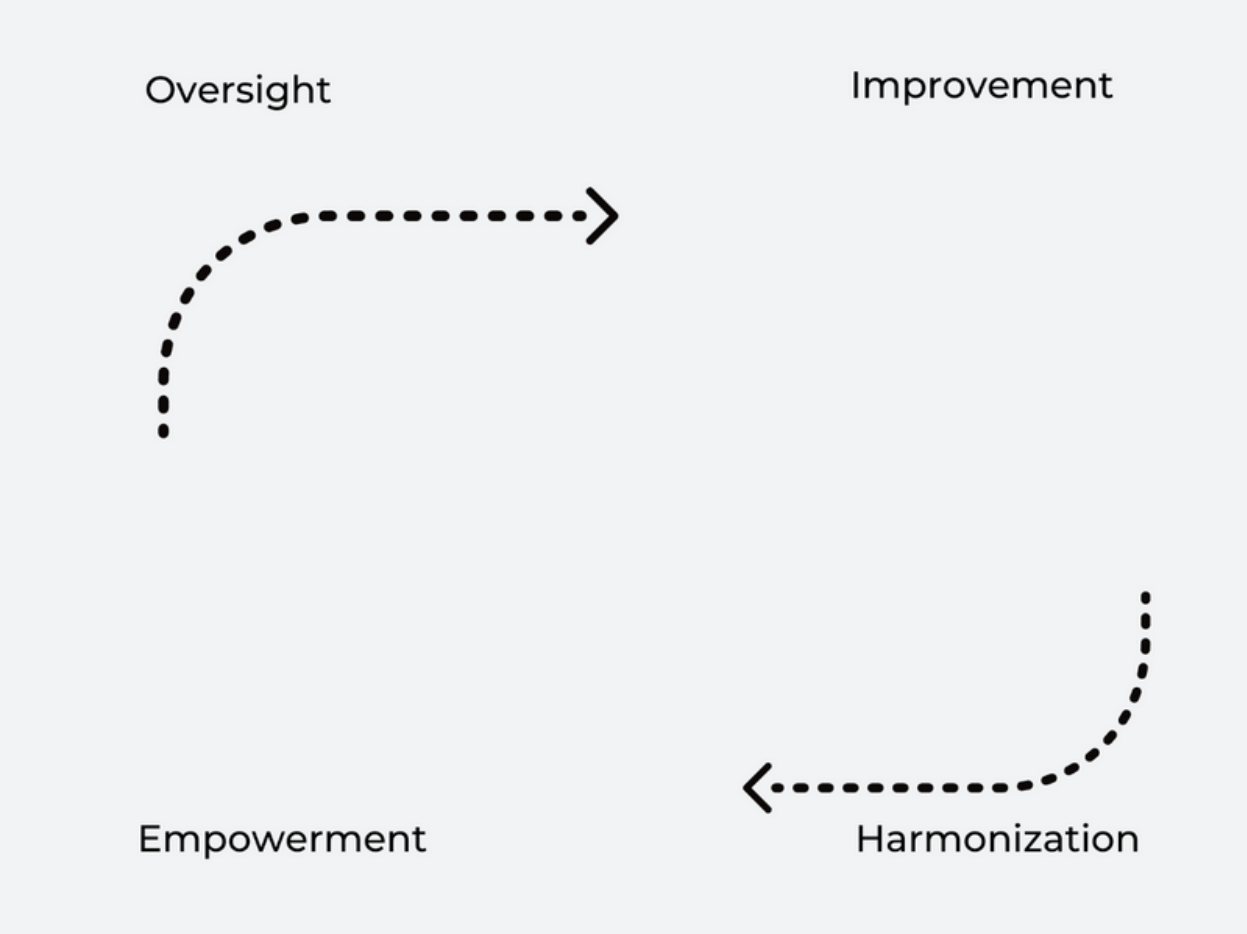
First, supervision ensures that services meet the required quality and ethical standards. By overseeing the work of less experienced individuals and third-party affiliates, supervisors can detect and correct errors, preventing potential harm to clients and the public. This is particularly important in healthcare, law, and finance, where even small errors can have significant consequences.

Second, supervision promotes adherence to specific legal and ethical guidelines. Supervisors must be well-versed in the laws and ethical codes relevant to their profession to provide effective guidance. This protects clients and the public from unethical and illegal practices and upholds the reputation and integrity of the profession.

Finally, through their role in professional development, supervisors help ensure those they supervise are competent and capable. By providing guidance, feedback, and mentorship, supervisors contribute to improving professional practices. This, in turn, benefits clients and the public, who can trust they receive services from skilled and knowledgeable professionals.

Four Supervisory Functions

Considering the specific challenge of supervising professionals using AI tools, let’s begin this discussion of a proposed framework by exploring the four supervisory functions: oversight, improvement, harmonization, and empowerment.



Oversight

Oversight involves overseeing the operation and deployment of AI tools in a professional setting. For instance, lawyers might utilize AI tools for legal research, document review, contract analysis, and predicting case outcomes. Proper oversight includes setting parameters for AI use and ensuring AI is leveraged properly for tasks it can effectively perform while using traditional methods when human judgment is essential. Tracking the effectiveness of AI and assessing its reliability and accuracy in producing consistent and legally sound results are also crucial oversight tasks.

Identifying areas that require human judgment is not unique to the legal profession. For example, physicians might utilize AI tools for diagnostics, patient data analysis, treatment planning, and predicting disease progression, but they also must deploy conventional methods when human insight is indispensable. Similarly, marketers might use AI tools for customer segmentation, predictive analysis, content creation, and forecasting market trends, but they still use traditional methods when human creativity is essential.

Managing ethical considerations related to AI systems focuses on adhering to principles that benefit everyone, including balance, fairness, and safety for all. In the legal context, managing ethical considerations might

involve protecting data privacy and confidentiality, and ensuring AI operations are not discriminatory. For instance, legal professionals using AI tools for contract analysis must ensure those tools are not inadvertently biased against certain contractual terms due to training data, potentially leading to unfair outcomes.

Improvement

The improvement function in AI supervision focuses on the continual evolution and improvement of the AI system. This requires a close working relationship between professionals (like lawyers) and AI developers or data scientists. Based on feedback from professional users, the AI system might need refinements to its algorithms or additional training with new data sets to enhance its performance and keep it up to date with the latest legal standards and practices.

For instance, an AI system used for legal research or case prediction might need to be updated as new legal precedents are set to maintain its accuracy and relevance. Similarly, if a law firm starts work in a new area of law, its AI tools may need to be trained with legal documents and cases related to the new practice area.

Empowerment

The empowerment function ensures the AI system has the necessary resources for its optimal operation and assists its users. This might involve creating a technical support team to address issues or glitches that arise while using the AI tool. Empowerment also ensures the necessary computational resources, data, and software are available and in good working order for the AI system to operate efficiently.

Empowerment might also involve ongoing training for professionals on how to use the AI system effectively. For example, a law firm might conduct regular training sessions for its lawyers on using its AI document review tool, including guidance on interpreting its outputs and troubleshooting common issues. In healthcare, a hospital might conduct regular training sessions for medical staff on using an AI diagnostic tool and provide guidance on interpreting outputs, integrating the use of AI into patient care, and troubleshooting common issues. In both contexts, the goal is to ensure that professionals can effectively and confidently use AI tools in their practice, ultimately improving patient and client services.

Harmonization

Harmonization plays a vital role in bridging the gap between the AI system and its human users, as well as other stakeholders. This function is essential for ensuring the safety and well-being of the individuals and communities served by the profession.

Consider city planners using an AI system for urban development. They will likely need to communicate and interpret AI-generated development plans to city officials, residents, and community groups. They provide a human perspective to the AI’s output, translating complex processes and results into understandable terms.

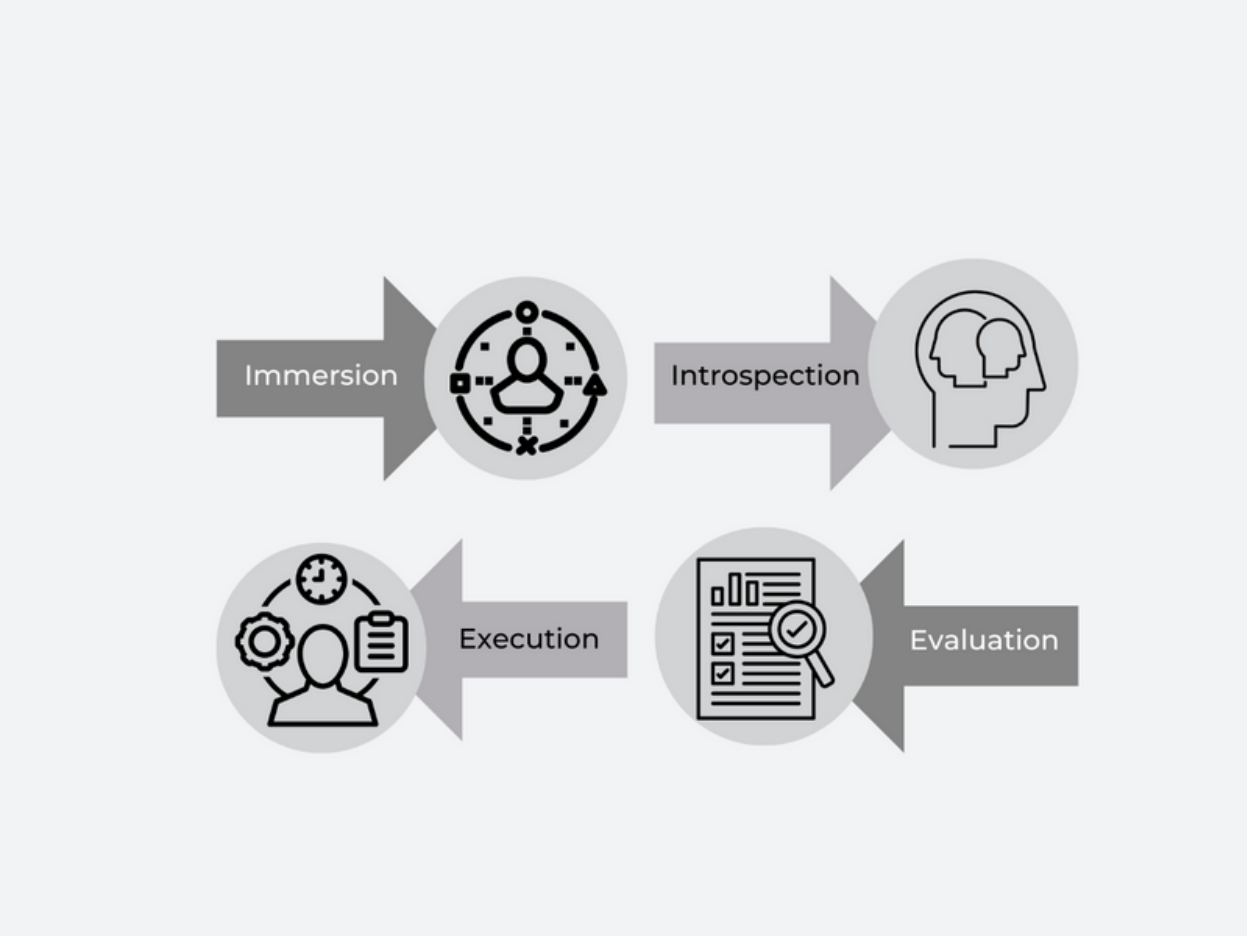
This harmonization function is particularly important when the AI system's recommendations must be adapted to the unique needs and constraints of the community.

The four supervisory functions ensure that AI tools are used effectively, ethically, and productively in professional settings. They enable professionals to leverage the capabilities of AI, while maintaining necessary human oversight and judgment. This balance ensures the safety and well-being of communities and the environment, leading to improved efficiency, accuracy, and outcomes.

The Four Steps of Supervisory Flow

Now, let's expand our understanding of the proposed supervisory framework for deploying AI tools by considering supervision in a temporal context, as a series of complementary functions that are implemented cyclically and iteratively over time. This is where the four steps of supervisory flow – immersion, introspection, evaluation, and execution – come into play.

The Four Pillars of Oversight are foundational in outlining the roles and responsibilities professionals must adopt when overseeing AI tools. The Four Steps of Supervisory Flow delineate the iterative process professionals undergo, from hands-on experience with AI to reflective actions for improvement. While the functions provide the structural framework for AI oversight, the stages offer a dynamic roadmap, guiding professionals through a continuous journey of AI mastery, ensuring ethical compliance and operational excellence.



Immersion

Immersion is the initial phase of the supervisory cycle, where professionals actively use an AI tool in their work. In the legal profession, lawyers might use an AI system for legal research, sifting through large volumes of legal documents, or perform contract analysis tasks. This direct, hands-on interaction allows lawyers to gain practical experience with the technology. They get to understand the workings of the AI system, including its strengths, weaknesses, and the various nuances of its operation. They observe its performance, the speed at which it completes tasks, and its effectiveness compared to traditional methods. At this stage, the lawyer, or other professional, might also encounter potential challenges, such as difficulty in interpreting the AI's output or concerns about its reliability.

In the field of data science, analysts might use an AI system for processing and analyzing large datasets, identifying patterns, or generating insights. In marketing, professionals might use an AI system for customer segmentation, personalized advertising campaigns, or social media analysis. In both cases, direct, hands-on interaction allows users to gain practical experience with the technology and familiarizes them with potential challenges.

Introspection

Once professionals gain sufficient experience with the AI tool, they move to the introspection stage, during which they think deeply about their experiences. Here, they consider how the AI system has impacted their workflows, any problems encountered, their initial reactions to these issues, the outcomes, and results for the client. They might also ponder their successes and overall impressions of the AI tool.

A lawyer may notice that the AI system greatly increased their efficiency in contract review, but they still needed to double-check outputs due to concerns about accuracy. A software engineer might realize that though the code analysis tool significantly improved their code quality, they still manually reviewed the changes due to concerns about false positives. A teacher might conclude that a generative AI tool enhanced student engagement through interactive activities, but it was still necessary to supplement the tool with additional resources to ensure a comprehensive learning experience. All three professionals might also reflect on their initial skepticism toward using technology and how that perception has evolved over time as they've witnessed positive results.

Evaluation

Following introspection, professionals enter the evaluation stage. Here, they critically examine their experiences and start identifying patterns, drawing out key insights, and making connections to broader legal concepts or principles. The professional may uncover the underlying causes of the issues they faced, potential improvements, and training needs.

For instance, a lawyer trying to understand the occasional inaccuracies of an AI system may discover that the inaccuracies occurred because the system was trained on a different jurisdiction's contracts, requirements, or regulations. This could indicate a need for more education about the system's inputs and its underlying algorithms.

Execution

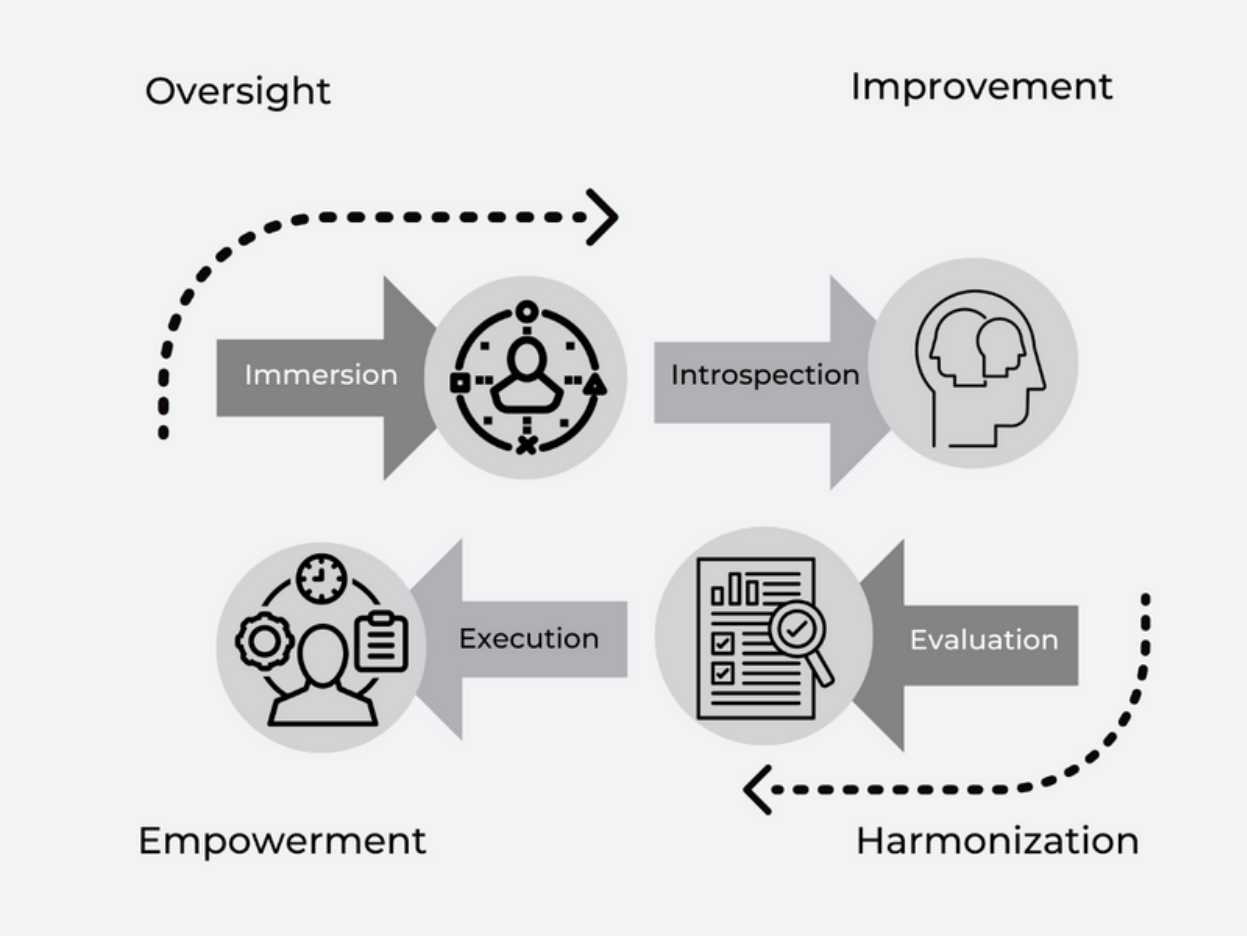
The final phase of the supervisory cycle is the execution stage. Based on their introspections and evaluations, the lawyer or other professional now takes appropriate steps to improve their use of the AI system. This can involve changing how they use the AI tool, seeking additional training, or collaborating with the AI developers to enhance the tool's performance.

The professional then might decide to attend a training session on the AI system to better understand its workings. They could also work directly with the AI developers to train the system on more relevant examples, improving its accuracy in their specific use case. Similarly, an educator using an AI system for personalized learning, for instance, might attend a training session and work with the AI developers to train the system on more diverse learning styles.

After the execution stage, the supervisory cycle begins anew, as the professional gains new experiences based on their most recent actions. Embracing this continual process of immersion, introspection, evaluation, and execution helps them adapt and evolve their use of AI tools, improving efficiency and effectiveness in their work, while also addressing any ethical, legal, or practical concerns that arise. This cycle is a dynamic process that supports continuous learning and professional development, ensuring lawyers or other professionals remain effective and ethical in using AI technologies.

How the Four Pillars of Oversight Align with the Four Steps of Supervisory Flow

Now, let's look at some additional examples to elaborate on how the four pillars of oversight align with the four steps of supervisory flow.



Oversight and Immersion

The oversight function involves a professional actively overseeing the deployment and utilization of AI tools in their practice. Consider a financial advisor using an AI-powered investment analysis tool. The financial advisor

would manage the tool's usage, establishing guidelines for its operation and monitoring its application in different tasks. This may involve deciding when and how the tool should be used, determining which investment analysis tasks it is most suited for, and ensuring it is used ethically and appropriately.

At the same time, a financial advisor managing the AI tool also gains direct experience, understanding the tool's capabilities, observing its performance in various tasks, identifying potential issues, and evaluating its effectiveness compared to traditional investment analysis methods. These first-hand experiences with the AI tool form the foundation for the subsequent introspection, evaluation, and execution stages.

The oversight function, paired with actual immersion, plays a crucial role in ensuring the benefits of AI tools are realized and maximized for clients, leading to more accurate and comprehensive investment analyses. This can result in better investment decisions, improved financial outcomes, and, ultimately, increased financial security.

Improvement and Introspection

The improvement function corresponds closely with the introspection stage of the supervisory cycle. After a lawyer or other professional has gained a certain level of experience with the AI tool, they engage in a process of introspection where they may consider the AI tool's strengths and weaknesses, the benefits and challenges encountered in using the tool, and its overall impact on their workflow and efficiency.

This introspection can then further inform improvement of the AI tool. For instance, if the lawyer determines the AI tool sometimes struggles with interpreting specific complex legal statutes, they could collaborate with the tool's developers to improve this aspect of its capabilities. In the interim, they could also develop new guidelines or procedures to address this limitation.

Empowerment and Evaluation

The empowerment function aligns with the evaluation stage of the supervisory cycle. After reflecting on their experiences, the professional more rigorously and critically analyzes and examines patterns in the AI tool's performance, identifying recurring issues or challenges, and extracting key learning points.

This evaluation helps the professional determine what is needed. For instance, if the lawyer or other professional determines that their colleagues often struggle to interpret the AI tool's output, they could organize training sessions to address that challenge. They could also liaise with the AI tool's developers to request enhancements to its user interface, making it more intuitive and user-friendly.

Harmonization and Execution

The harmonization function relates to the execution stage of the supervisory cycle. Based on their introspection and evaluation, the professional decides on appropriate execution. To ensure a comprehensive decision-making

process, it is important to consider the potential risks, benefits, and feasibility of each action, while also keeping in mind legal and ethical considerations. The professional may need to mediate between stakeholder groups, such as colleagues using the AI tool, the tool's developers, and the clients who benefit from the tool's output. By taking a systematic approach and involving all relevant stakeholders, they can make well-informed decisions that promote fairness, accountability, and positive outcomes.

For example, the lawyer or other professional may decide that a certain AI tool isn't suitable for certain legal cases due to the complexity of relevant laws. In this case, they would need to mediate between the developers and the legal team, explaining why the tool isn't suitable for these cases and exploring potential solutions. This could involve working with the developers to enhance the tool's capabilities or providing guidance to the legal team on when and how to use the tool effectively.

Conclusion

The importance of AI supervision cannot be overstated. The proposed model for integrating supervisory functions and cycles is intended to provide professionals grappling with the implications of intelligent tools with a logical framework for oversight and adaptation. A supervisory framework forms the backbone of quality control, ensuring that AI systems function as intended and deliver accurate and useful results. Individual supervisors monitor the use and performance of AI, identifying anomalies or errors. When issues arise, they step in, taking corrective action to maintain the integrity and reliability of the system.

But the role of supervision extends beyond mere functionality. Supervisors are the gatekeepers of ethical compliance – a crucial consideration in the era of AI. As AI systems become more integrated into our professional and personal lives, our use of them must adhere to ethical standards and legal regulations. Supervisors ensure that intelligent systems do not overstep their boundaries or violate principles that range from data privacy laws to guidelines against bias and discrimination. In essence, supervision is the compass that guides the AI ship, ensuring it stays on course in terms of performance and ethical conduct. It is the key to unlocking the full potential of AI while safeguarding the interests of all stakeholders.

This article includes content that has been submitted and accepted for the media section of law.mit.edu, but is in the process of being edited. Similar to the publication of pre-prints, the purpose of this page is to get content out in the world so that they may receive more feedback before final publication. Accordingly, this page is subject to revision and fluctuation. To comment, simply create a PubPub account, and sign in. As you're reading, you can highlight the text you want to comment on, and an icon will pop up for you to write your comment inline. You can also make general comments at the end of each piece of content.