29 Aligning GenAI with Organizational Values

For all the noteworthy breakthroughs in Generative AI since late 2022, there have been just as many reports of ethical concerns, biases, and potential job impacts. Many are hyperbolic; but many others are real.

CoE members or business executives - anyone with high Emotional IQ – should read and then consider this chapter. Your organization's culture and tolerance for risk will guide decisions of how quickly and broadly to address the human risks of AI. Prudent executives will be out in front of them.

This chapter starts with an overview of Microsoft's role in safe, responsible AI, then outlines where organizations step in to chart their own course.

How Microsoft Supports Responsible AI

Out of all the options in the public market, Microsoft 365 Copilot is one of the safest ways to harness generative AI in an enterprise. There are at least some administrative controls, and Microsoft manages the service with a service description, copyright protection, and a commitment to cooperating with governments and industry consortiums to set and follow standards.

Microsoft has invested in a cross-company program to ensure that its AI systems are responsible by design. The company launched a committee in 2017 to focus on responsible AI issues and help craft the AI principles that it adopted in 2018.[1] Microsoft has publicly shared its Responsible AI Standard, a framework to guide how the company builds AI systems. This standard sets out Microsoft's best thinking on how to build AI systems that uphold its values and earn society's trust.

Microsoft's responsible AI principles include:

- Fairness
- Reliability and safety
- Privacy and security
- Inclusiveness
- Transparency
- Accountability[2]

Despite Microsoft's philosophies, principles, and best efforts, there's no overcoming the randomness of how LLMs work. LLMs inherently 'predict' what should come next based on massive amounts of training data. The path it takes from Alpha to Omega is indiscriminate and unpredictable. Mistakes will be made. Humans may not like the results.

Making Your Own Statement

Prudent organizations will recognize LLMs' randomness and build some of its own policies and principles. In lieu of a better approach, set your own intentions on the same six topics above, and consider them when writing the organization's AI Statement and AI Policy.

Fairness

Make known your commitment to ensure that there is no intentional bias or discrimination in the AI systems you buy (and especially those you build).

Reliability and safety

Commit to ensuring that the AI systems you use are reliable and safe, and that they operate in a manner that is consistent with the expectations of users and society. To be safe, your policy could be to deny the use of any unsanctioned AI (shadow IT and the like).

Privacy and security

Make your commitment known about the respect for the privacy and security of your employees' data, and that you operate in a manner that is consistent with users' expectations and legal requirements. Like cybersecurity in general, these days, this is not just an IT or Chief Security Officer function.

Inclusiveness

Set your course on ensuring that AI systems that you use are inclusive and consider the needs of all users.

NOTE: This principle infers that everyone can use AI. That may conflict with a CFO's budget constraints, since Copilot's $30/month fee often makes it untenable to use across an entire organization.

CFO Tip: For people where $30/month for M365 Copilot isn't justified, consider offering Microsoft 365 Copilot (Basic) People can interact with a ChatGPT-like interface, but instead of using the public models, Copilot provides protects prompts and doesn't train the public model.

Now, before reading last two principles, realize they are less straightforward, and are primarily aimed more at the developers of AI. For most enterprises and Copilot users, they're not pertinent.

Transparency

This principle states that AI systems will be transparent, and that users are able to understand how the systems operate and make decisions. **However, the reality is that LLMs are not always predictable, and that (most) users won't have the time (nor interest) in understanding how they work.**

Accountability

This principle states that an organization can take responsibility for the actions of its AI systems. **This is difficult in a cloud environment, since the models and their innerworkings are largely out of the ordinary organization's control.**

Putting These Principles into Practice

These can be very esoteric terms and principles. How does a leadership team or CoE put them into practice?

The table below connects common Responsible AI principles with practical controls organizations can implement. Use it as a starting point for CoE members and stakeholders about how you want AI to take shape in your organization.

Principle	What It Might Mean to You	Example Control to Encourage or Enforce
Fairness	AI should not create unjustified bias or discrimination.	Human review of high-impact decisions.
Reliability & Safety	AI systems should perform consistently and as intended.	Testing before deployment and ongoing quality monitoring.
Privacy & Security	Data must be protected throughout its lifecycle.	Data classification, DLP, and access controls.
Inclusiveness	AI should be accessible and beneficial to diverse users.	Accessibility reviews and diverse stakeholder testing.
Transparency	Users should understand when and how AI is being used.	AI usage disclosures and documented decision processes.
Accountability	Humans remain responsible for outcomes.	Named business owner for AI's outputs, and approval for AI workflows.

And now… onto something else that's equally difficult to control, people!

The Role of Responsible Users

Remember, Copilot is like an electronic intern. People direct it in the manner of their choosing. They too have a responsibility to use AI in an ethical and transparent manner. This includes ensuring that AI systems are used in a way that respects human values, rights, and dignity, and helps to prevent or mitigate the potential harm of AI.

By understanding responsible AI practices, users can help to build trust in AI and ensure that the technology is used for the betterment of the organization, and society.

In any policy or principle that an organization publishes, encourage users to report inconsistencies with what they're seeing in a private, safe way. The reporting process should be similar to the way other policy violations or HR incidents are reported.

If someone calls the IT help desk to report incorrect, biased or offensive outputs, will they know what to say and where to escalate?

The Role of Leadership

Continuing reading here if you've not (yet) launched Copilot in a meaningful way, or if you'd like to assess how well you've covered some key soft success factors. It may be worth doubling back and addressing any missing steps.

Ensuring a smooth and responsible operation and expansion of Microsoft 365 Copilot is a multifaceted approach. To date, very few early adopters have begun ensuring their use of AI aligns with their goals and culture. Below find some suggested steps:

Educate transparently.

- **Announce internally.** Hold an all-hands meeting or company-wide email announcing Copilot and/or your AI Charter and CoE. Emphasize the potential benefits for efficiency and productivity in your organization.
- **Ensure inclusivity.** Even though the licensed version of Microsoft 365 Copilot may not be available to everyone in the organization, employees with a browser and a company login can use M365 Copilot Basic (formerly Bing Chat Enterprise) for a safe and secure work GenAI experience.
- **Offer training sessions.** Develop training sessions for different user groups. Explain what Copilot is, how it works, its limitations, and potential issues. This requires some regularity, and remember, you're not training about a tool or technology, but a different mindset.
- **Communicate openly.** Encourage open dialogue and address any concerns employees might have about job security, ethical implications, or privacy.
- **Publish an Acceptable AI Use Policy.** Include steps to report issues or policy violations. The org or each department should consider guiding its associates about whether or not to use AI for::
    - Employment decisions without human review
    - Performance reviews generated solely by AI

- Compensation recommendations
- Customer communications without appropriate oversight
- High-risk regulatory or legal decisions

Focus on Human *PLUS* AI Collaboration.

- **Highlight what's complementary.** Emphasize Copilot as a tool to augment human capabilities, not replace them.
- **Stress accountability.** Stress that while Copilot can recommend, humans must remain accountable. Who will take the blame when AI is incorrect? Reread Chapter 1 if you think your people will blame Microsoft or the LLMs.
- **Offer upskilling opportunities.** Provide training and resources to help employees develop skills that complement Copilot's capabilities, such as data analysis, critical thinking, and problem-solving. See Chapter 31 for more on this topic.

Address Ethical Concerns.

- **Support security.** Users should be aware that sensitive data may be exposed or produced by Copilot, and that they should handle situations according to the Acceptable Use Policy, and comply with relevant regulations.
- **Be aware of bias.** There have been few if any reported issues, as LLM's guardrails seem to care for this concern. Even still, educate managers and employees to identify potential biases in Copilot's suggestions and to use their judgment when working with them. Develop a process to report and address any concerns.

Monitor and Evaluate.

- **Gather metrics.** Organizations that are concerned about costs should make it known that usage will be monitored, token limits established, and licenses reassigned if there is low (or no) activity. Employees should commit to providing regular feedback and responses to surveys to gauge Copilot's impact. In the future, evaluating the human-review compliance rate, policy exception requests, and sensitive data exposure events.
- **Define a process for AI incident response.**
  - How can people safely report issues (especially about data overexposure) and who will appropriately handle?

Alignment with Organizational Goals.

- **Connect to Goals.** Employees should know how Copilot will contribute to achieving specific organizational goals, such as increased efficiency, improved communication, or higher-quality work.

- **Call upon the culture.** Encourage employees to use Copilot in a way that aligns with the company culture. If the focus is on innovation and human connection, Copilot should be used to enhance those aspects, not replace them.

Culture Will this Chapter for Breakfast

Peter Drucker would likely agree that an organization's culture will inherently guide its course to AI. This chapter was written for those who want to take prudent steps, but it's acknowledged that they're not **all** necessary to get started with Copilot.

After all, even the wary Max Tegmark agreed, Copilot is a rather low risk use of AI in the overall picture. But still, in these early days, the key to aligning with organizational values is open communication, user education, and a focus on Human *PLUS* AI collaboration. Implementing policies and practices to that end will maximize the benefits of this technology in a responsible way.

References and Resources

[1] https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf
[2] https://servicetrust.microsoft.com/DocumentPage/a0682452-d305-4abf-be19-505c3e01c145