Microsoft

Power CAT

Power CAT AI Webinars

Building an AI-ready organization

Microsoft

Power CAT

Navigating AI risks

Imagine you work for a financial organization that is launching an AI-powered loan pre-approval system

The system predicts if loan applicants are ”high-earners” to pre-approve specific products.

Objectives

A deeper understanding of AI risks

Explore the types of harm AI can cause, from unfair decisions to reputational damage.

Practical ways to respond

Learn how responsible AI principles can help you identify, frame and respond to issues.

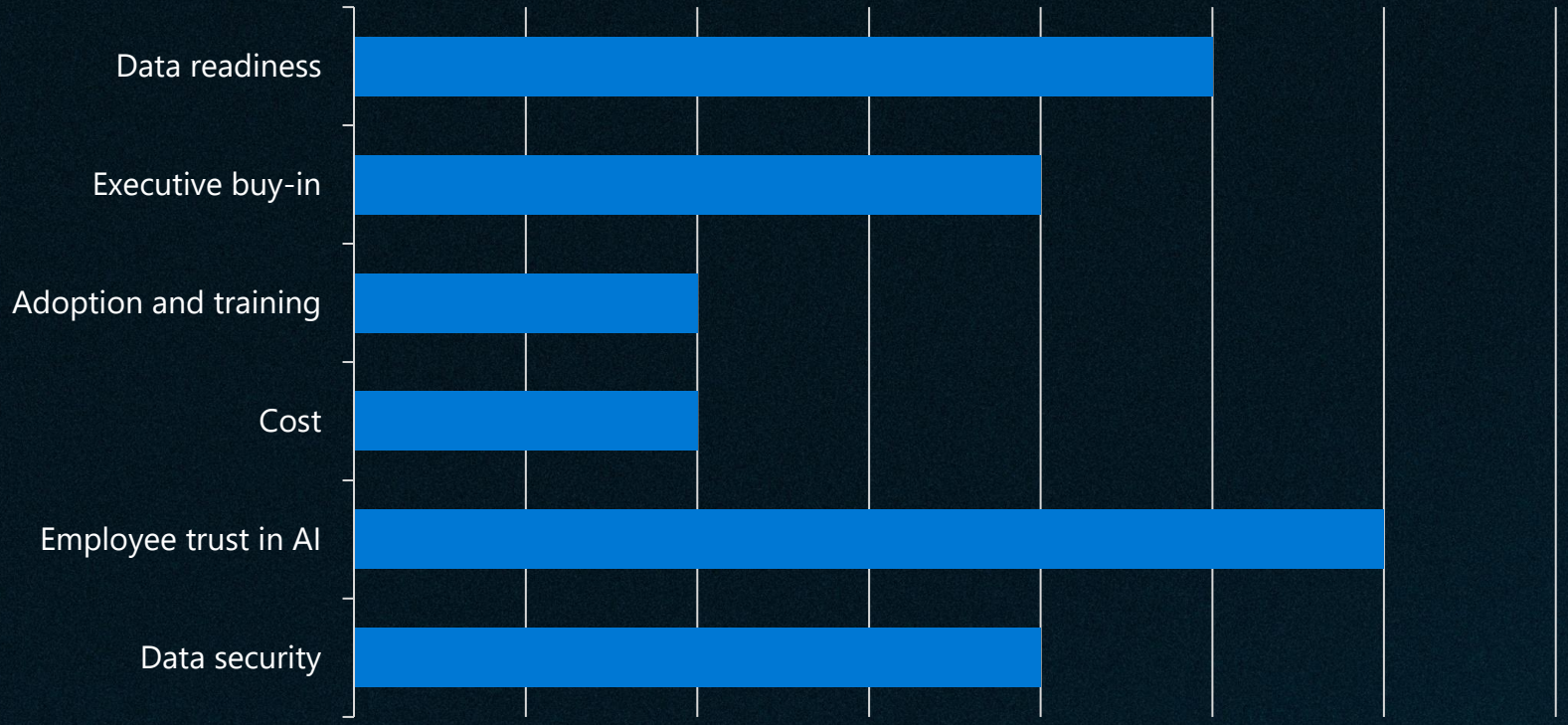
Walk away with a roadmap

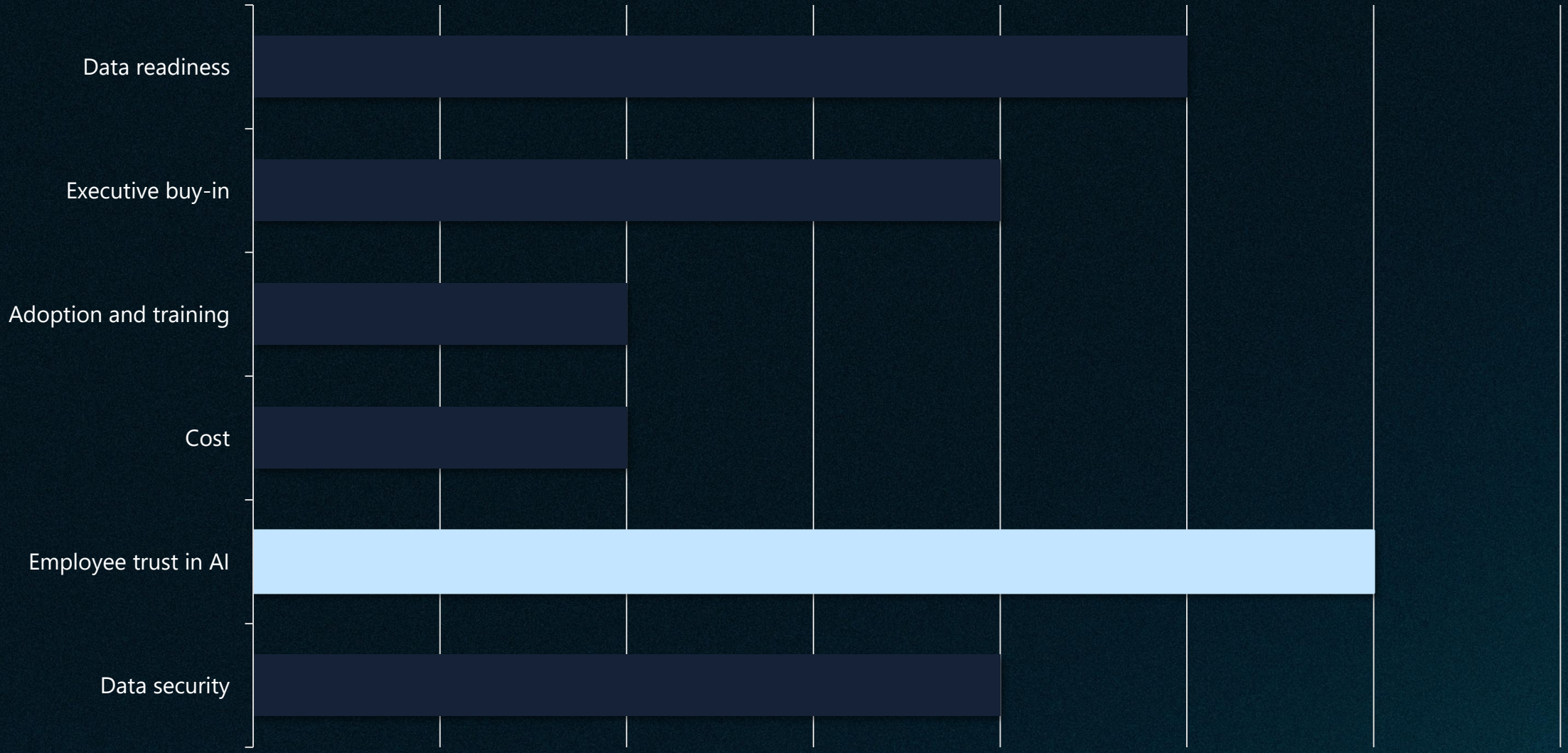
Use our card sorting activity to build an action plan that reflects where you are, and where you want to go next.

Understanding AI risks

What do you consider to be the single greatest risk associated with AI?

Understanding AI risks




Data readiness
Executive buy-in
Adoption and training
Cost
Employee trust in AI
Data security

Employees lack trust in AI because it makes decisions they can’t see, understand, or challenge, leaving them feeling powerless and suspicious.

Employees fear AI because they see it as an unpredictable force that could replace their jobs, judge their performance unfairly, or amplify hidden biases.

It might amplify hidden biases

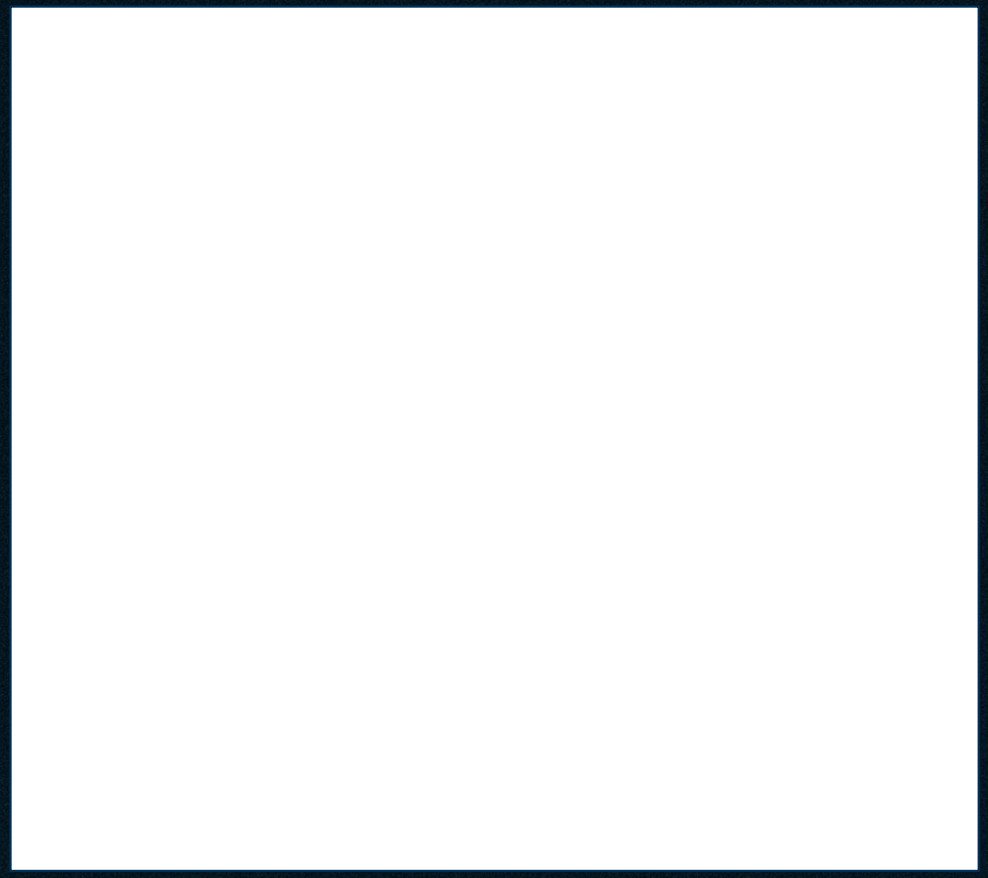
Decisions that can’t be seen, understood or challenged.

Fairness

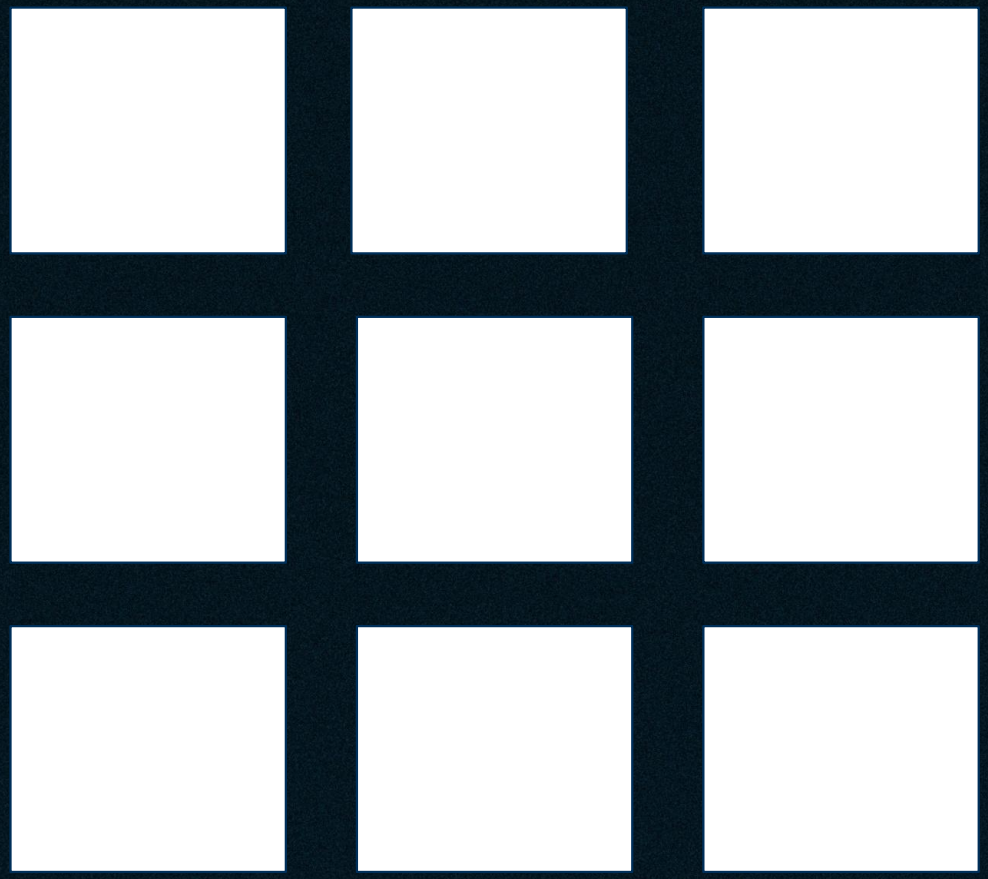
It might amplify hidden biases

Transparency and accountability

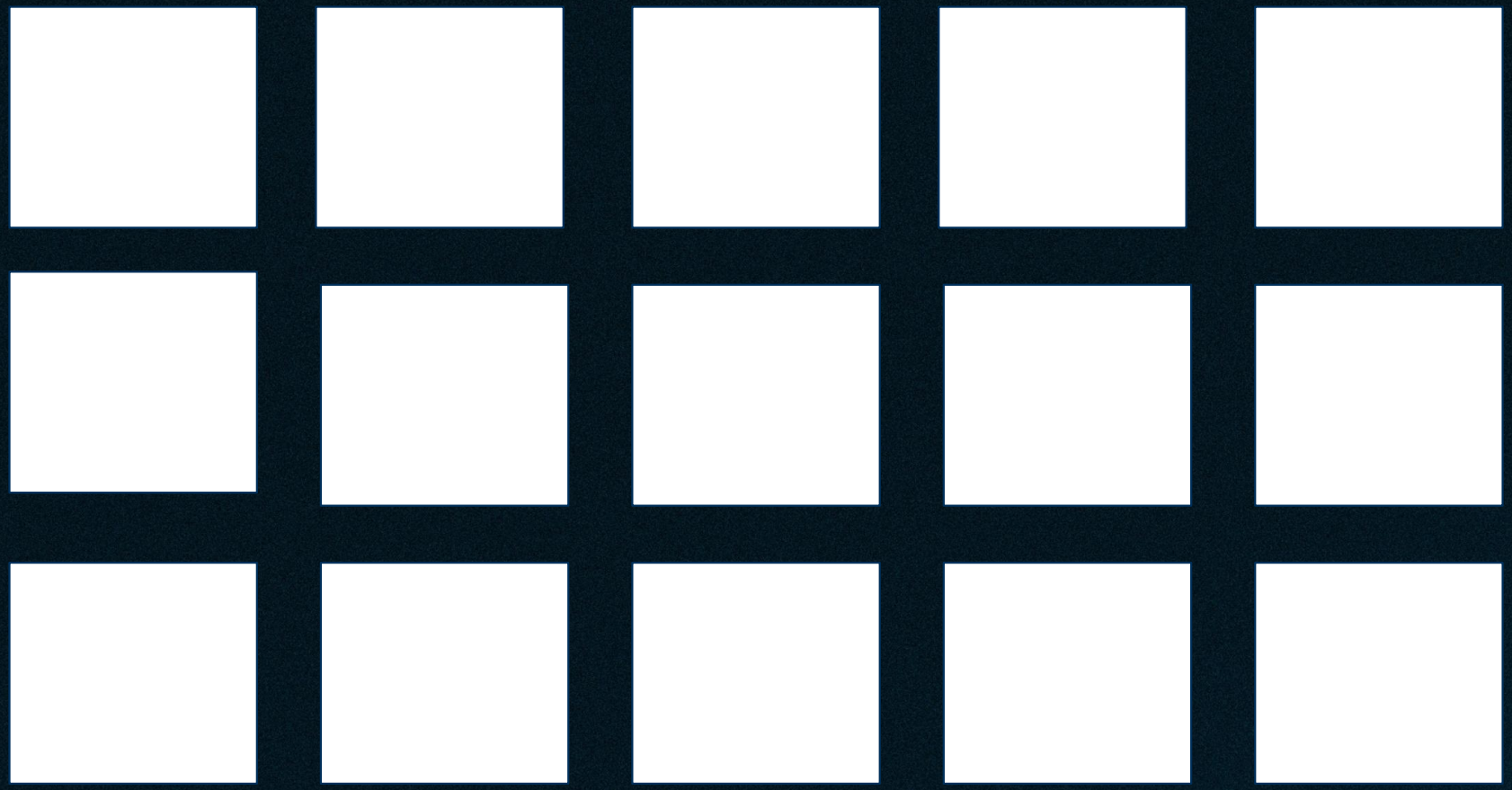
Decisions that can’t be seen, understood or challenged.



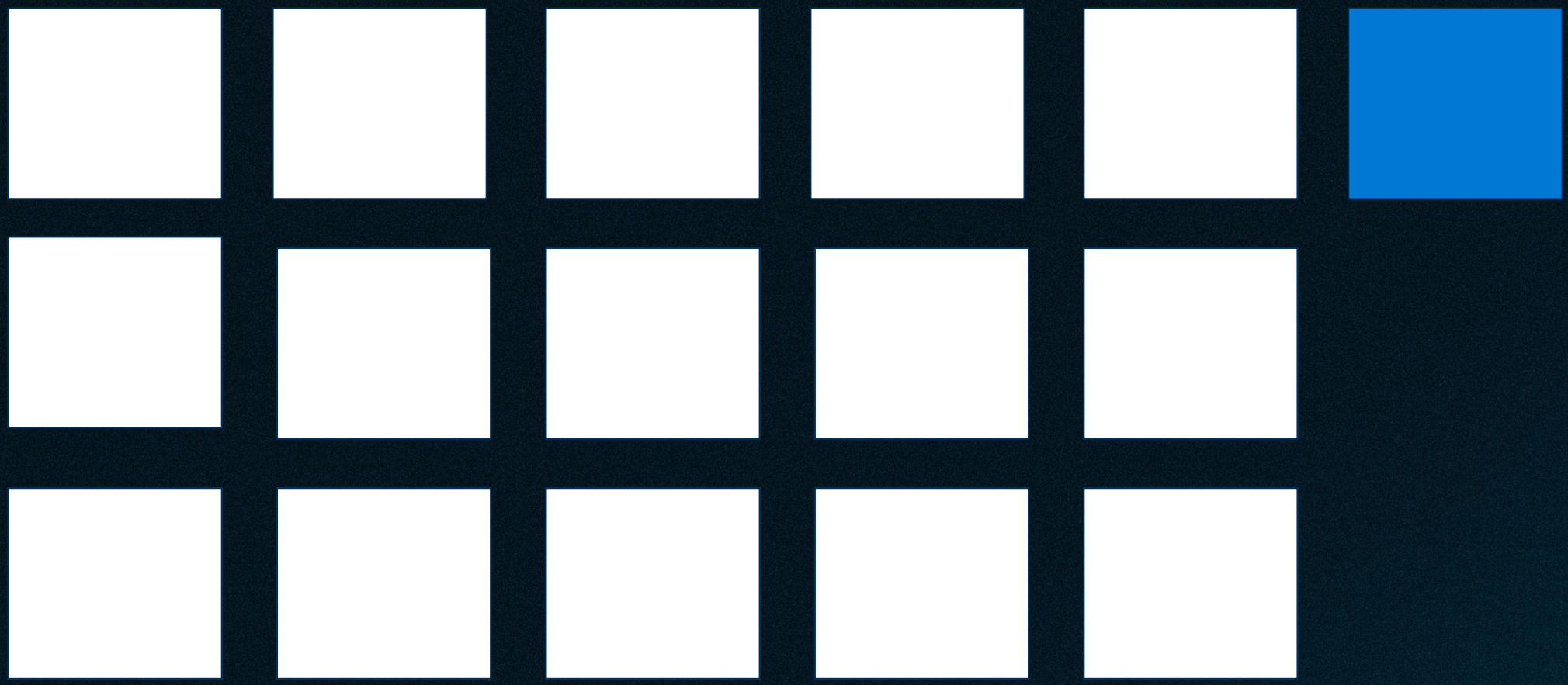
This is a square

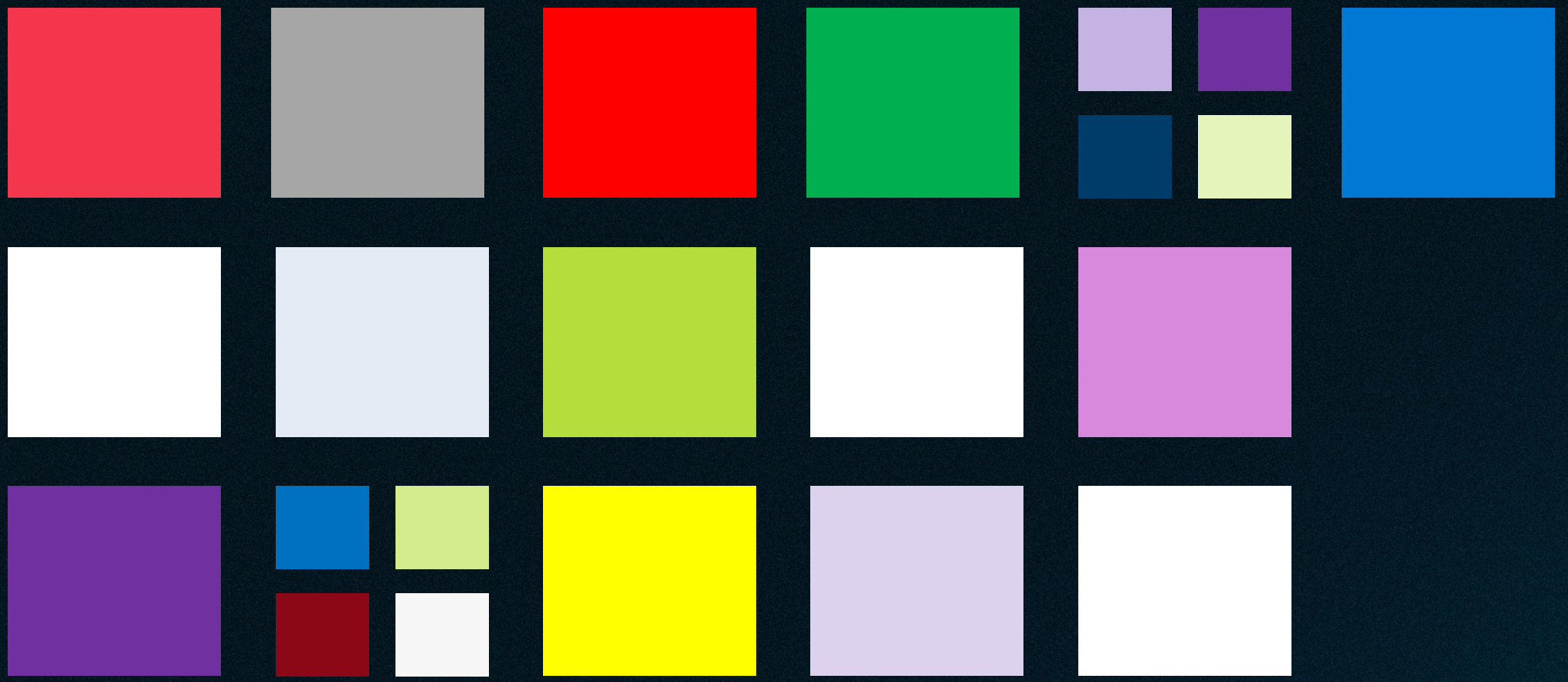


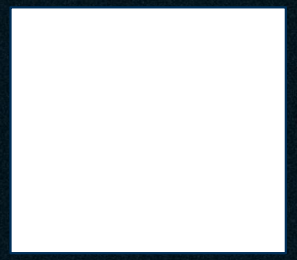
These are also squares



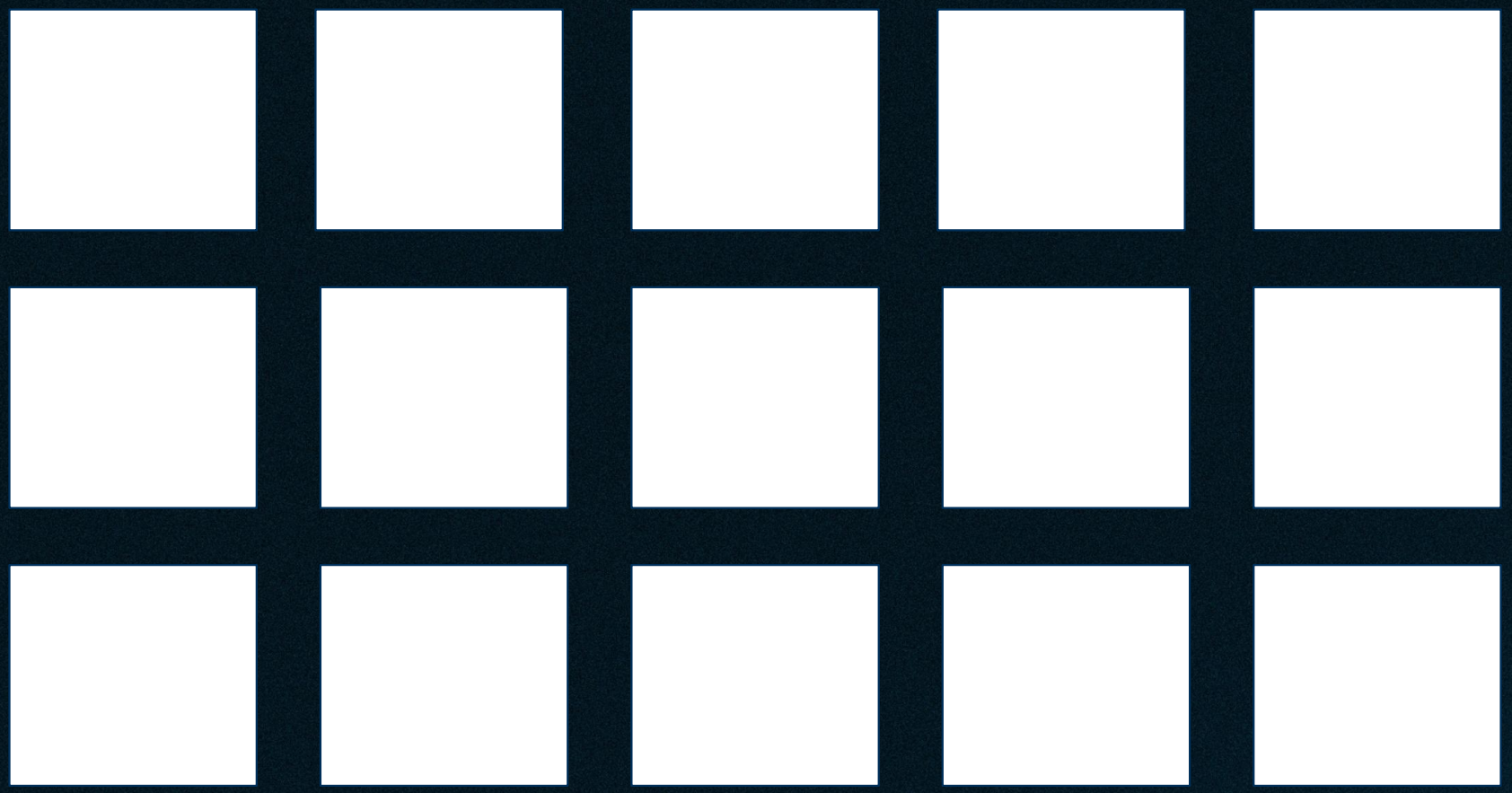
And so are these





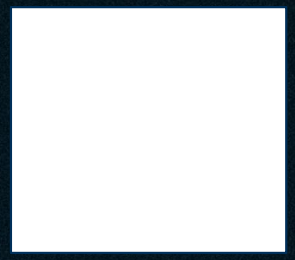


Squares come in all
different sizes and colors

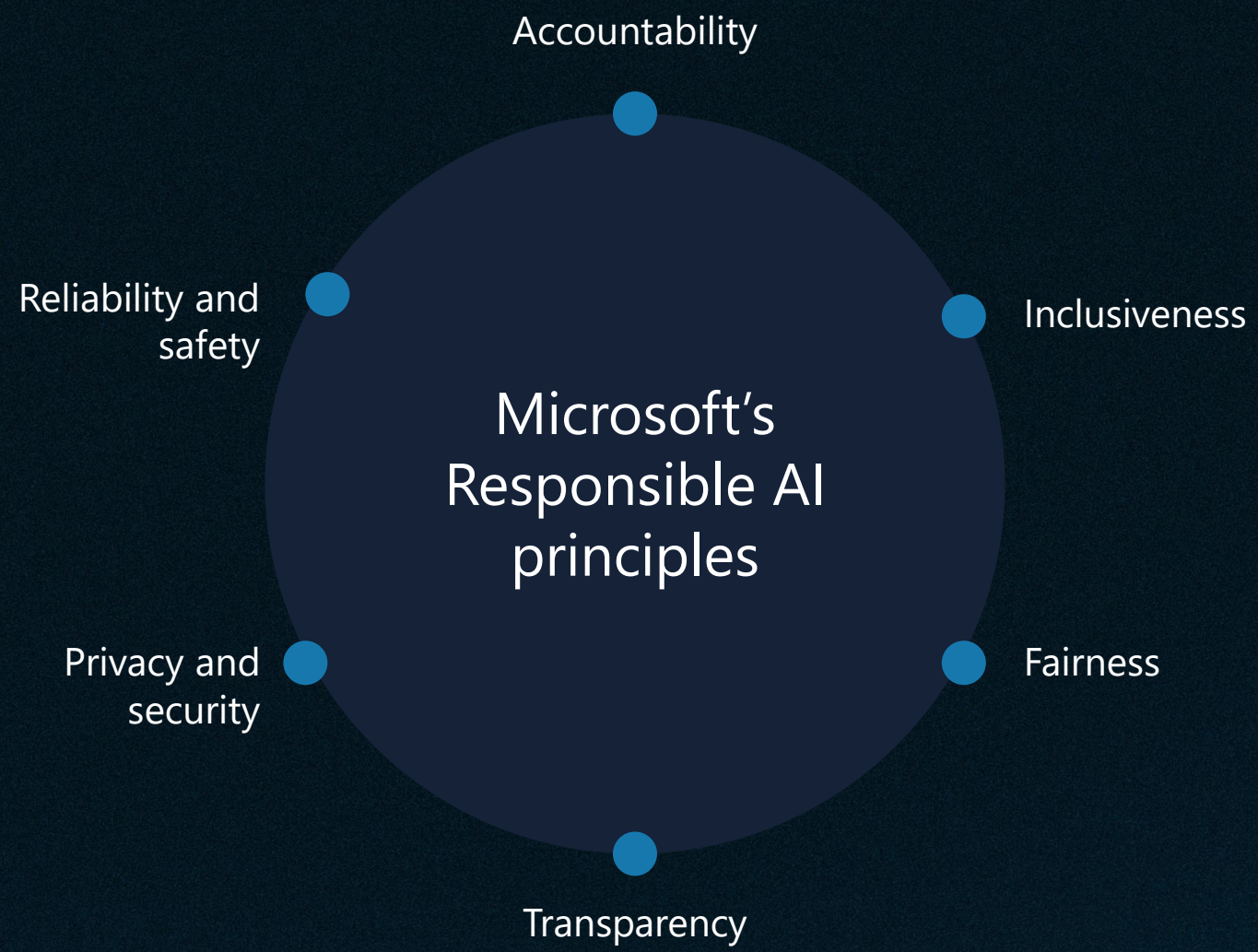


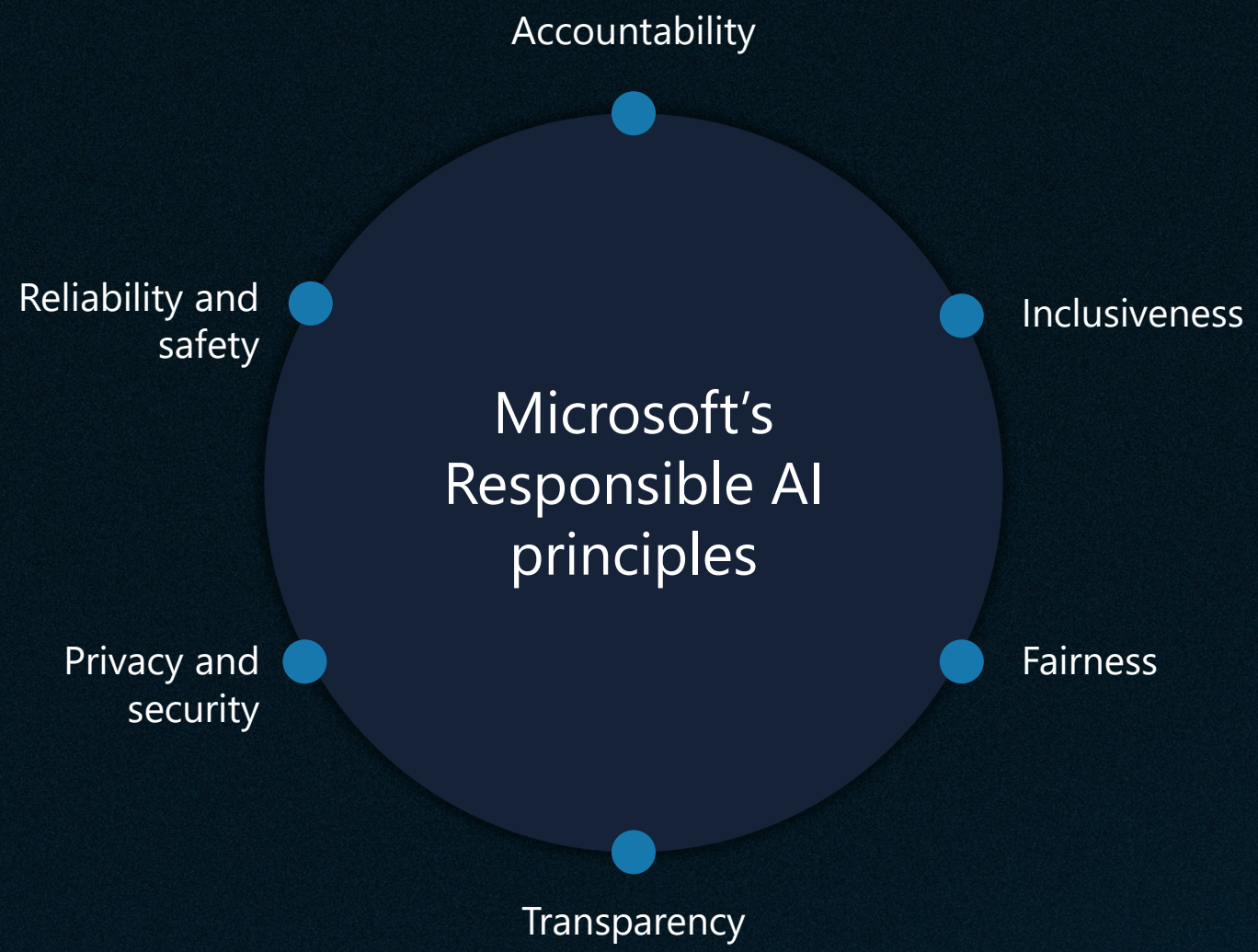
So if our models are trained on data that doesn’t accurately represent the real world…

Square



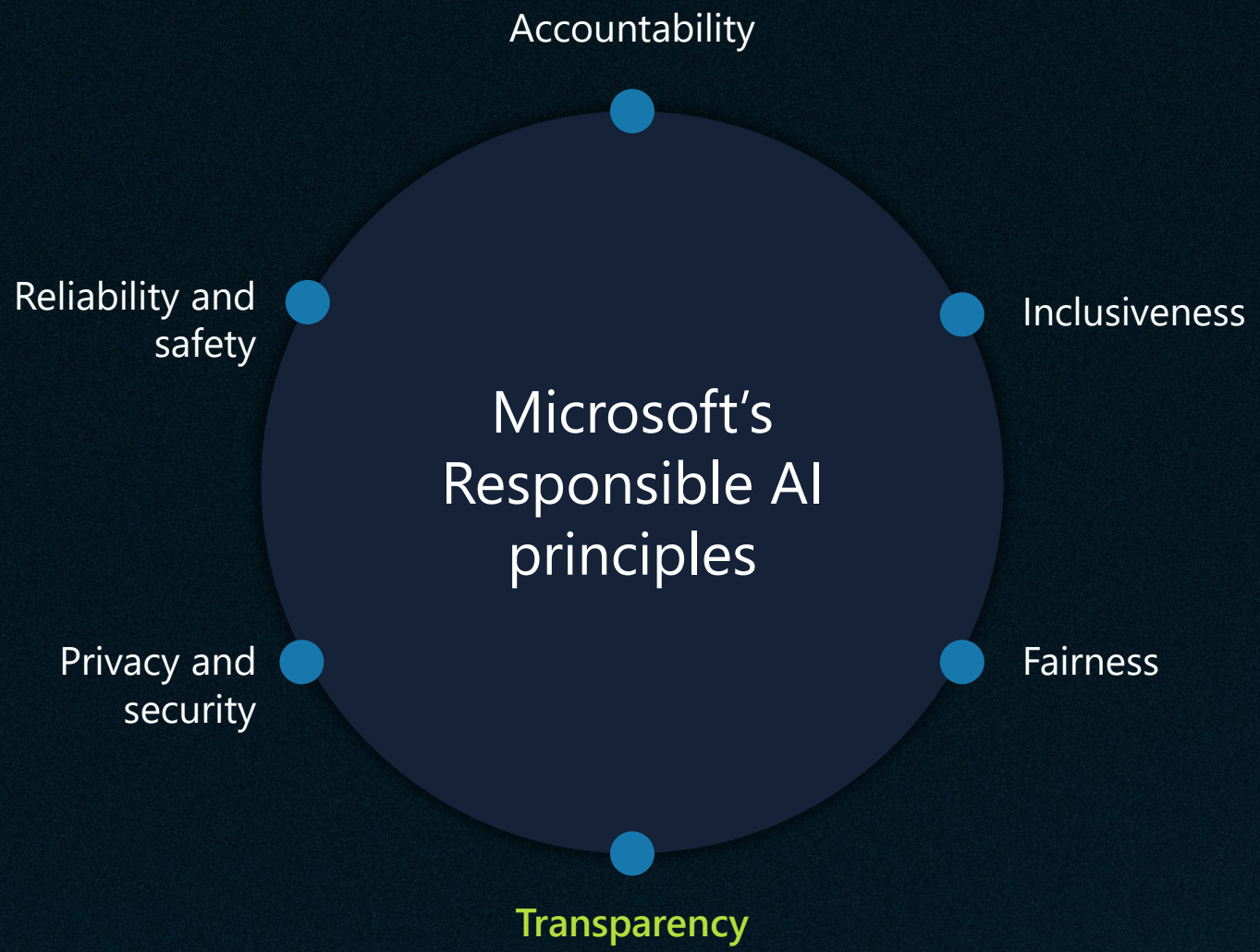
Not square


Accountability
Inclusiveness
Fairness
Transparency
Privacy and security
Reliability and safety
Microsoft's Responsible AI principles



Building blocks to enact principles

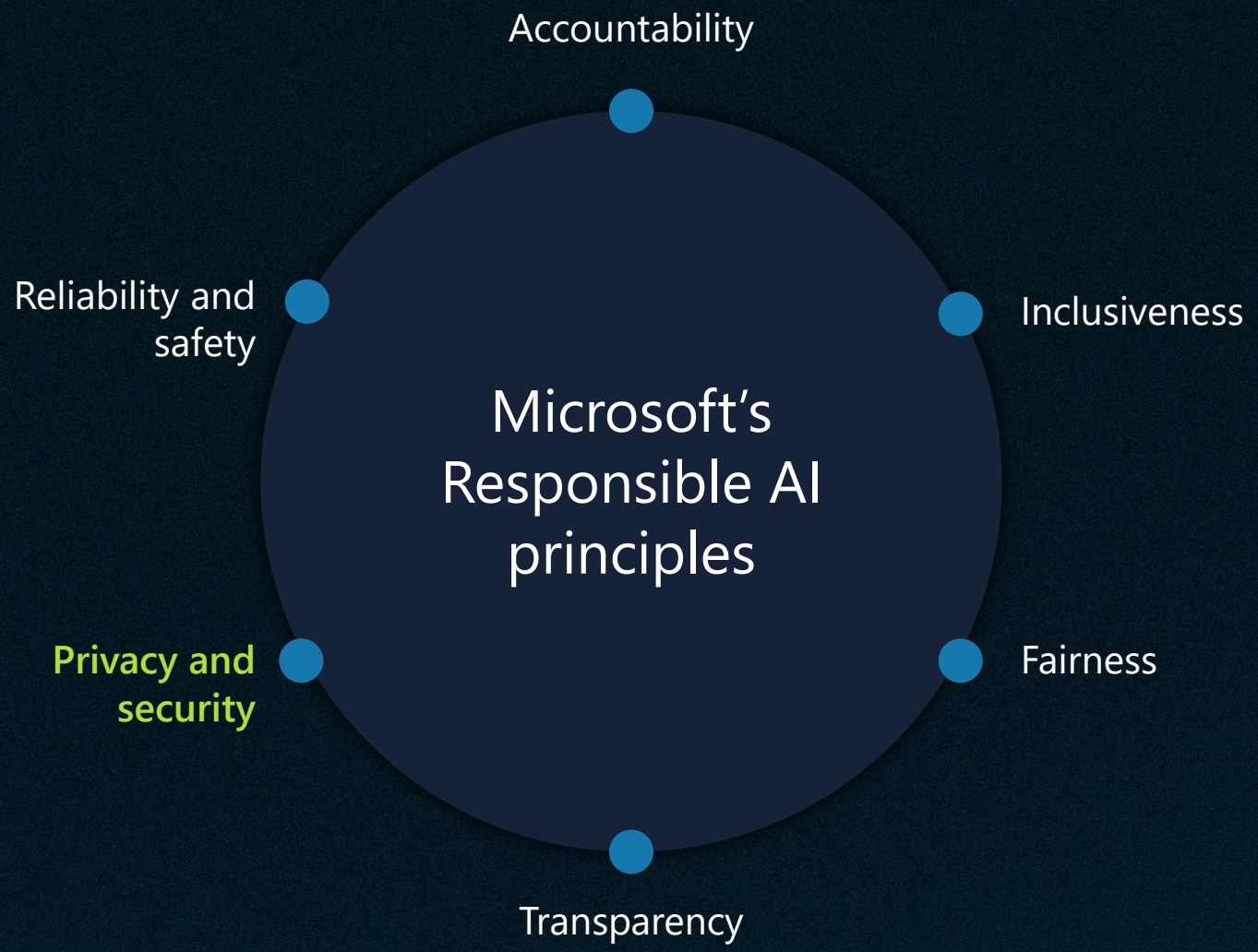
- Tools and processes
- Training and practices
- Rules
- Governance



Transparency

How might people misunderstand, misuse, or incorrectly estimate the capabilities of the system?

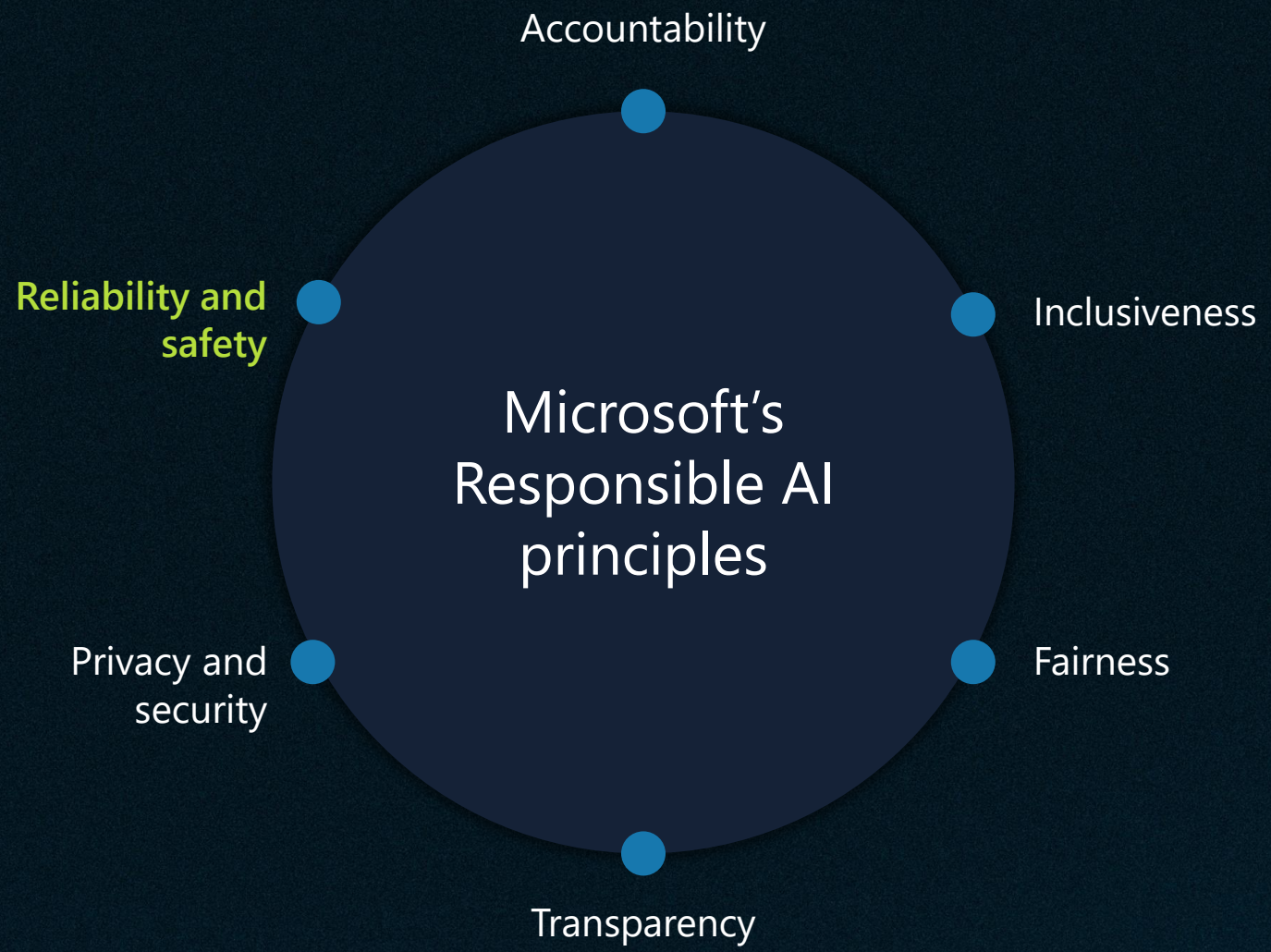
It means clearly explaining how AI systems make decisions, providing visibility into their operations, data sources and limitations.



Privacy and security

How might the system be designed to support privacy and security?

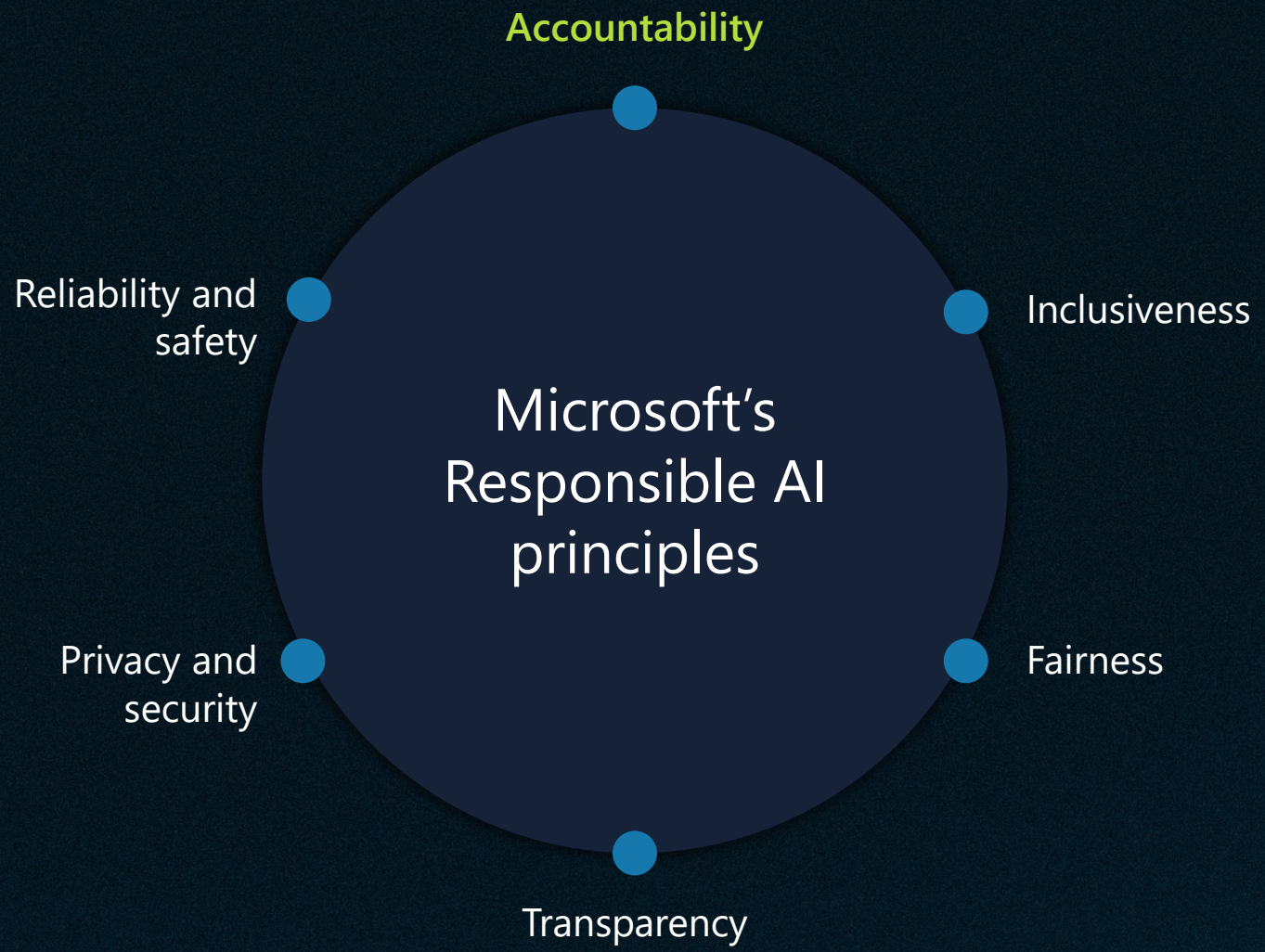
It involves protecting user data and ensuring AI systems are safeguarded against misuse or attacks.



Reliability and safety

How might the system function well for people across different use conditions and contexts, including ones it was not originally intended for?

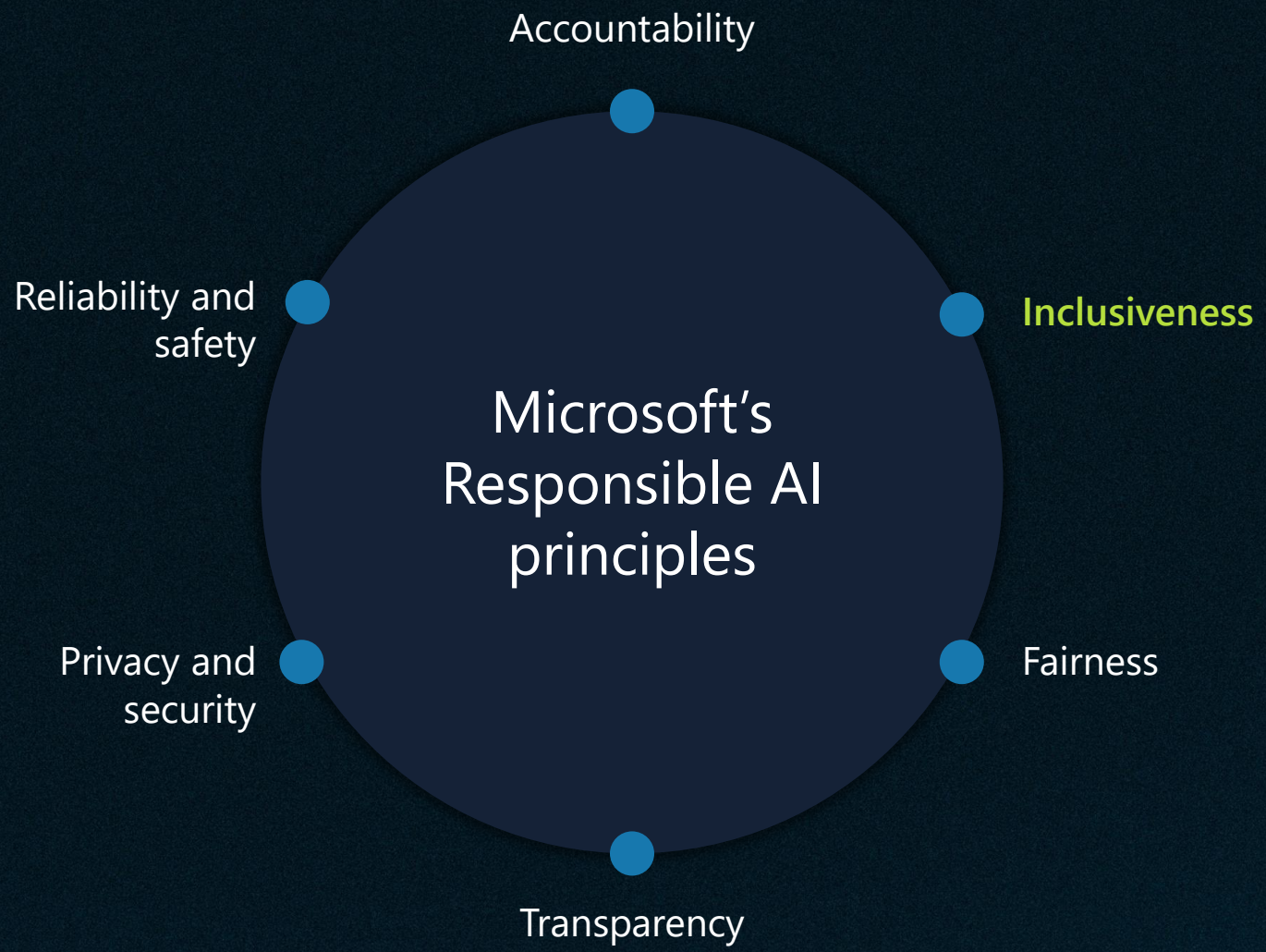
Consider how the system performs consistently, especially in critical situations. It involves preventing failures, managing errors responsibly and proactively minimizing harm.



Accountability

How can we create oversight so that humans can be accountable and in control?

People and organizations remain answerable for AI decisions and impacts – establishing mechanisms for oversight, and addressing issues transparently



Inclusiveness

How might the system be designed to be inclusive of people of all abilities?

It means designing systems that empower everyone equally, regardless of their background or abilities.

Fairness

How might an AI system allocate opportunities, resources, or information in ways that are fair to the humans who use it?

Systems must treat everyone equitably, without discrimination or bias.

Let’s go back to our scenario…

The system predicts if loan applicants are “high-earners” to offer specific products – results are not looking good

Customer services are taking numerous complaints about loan offers

An audit revealed loan approval rates were significantly lower for female applicants than male applicants.

What do you think might be causing
incorrect predictions?

How would your organization respond?

Panic and switch things off	Basic awareness and manual intervention	Investigate and learn	Proactive governance and transparency	Continuous oversight
The team shuts down the AI model in a panic and switches to manual loan approvals.	The team manually reviews recent loan decisions to identify unfair patterns.	The team retrains the model to balance approval rates.	The team investigates and explains model decisions (e.g., why a loan was denied).	Automated fairness checks run during model training and deployment.

It’s rarely just one thing…

Poor or unrepresentative training data	Lack of documentation	No testing or monitoring in production	Use of sensitive personal data without review
AI learns from what it’s given – and biased or incomplete data leads to skewed outcomes.	Key design choices aren’t written down, making it hard to explain or revisit decisions.	Even good models degrade over time. Without oversight, issues go unnoticed.	Using demographic data without safeguards raises ethical and legal risks.

The problem with problems…

…is that they don’t arrive with labels like ‘fairness issue’ or ‘transparency gap’. They just show up as complaints, errors or uncertainty

From reaction to response

Poor or unrepresentative training data

Lack of documentation

No testing or monitoring in production

Use of sensitive personal data without review

From reaction to response

Poor or unrepresentative training data	Lack of documentation	No testing or monitoring in production	Use of sensitive personal data without review
FAIRNESS	**TRANSPARENCY & ACCOUNTABILITY**	**RELIABILITY & SAFETY**	**PRIVACY & SECURITY**

Poor or unrepresentative training data	Lack of documentation	No testing or monitoring in production	Use of sensitive personal data without review
FAIRNESS	**TRANSPARENCY & ACCOUNTABILITY**	**RELIABILITY & SAFETY**	**PRIVACY & SECURITY**

Response plan

- Review data sources for representation gaps.
- Add real-world examples that reflect full customer base.
- Consider running fairness assessments before retraining models.

- Create a simple model summary that records how the model was built, what data it uses and known limitations.
- Make this visible to your product, compliance or legal teams to support reviews.

- Setup basic monitoring – even a monthly check – to review if outcomes are drifting.
- Add alerting for big drops in approval rates or large changes across demographic groups.

- Identify and flag sensitive features like age, race, gender that may need extra review.
- Use internal guidance (or create it) to help teams assess when sensitive data is used ethically and legally.

So… having response plans ready for each principle seems like a smart move, right?

Now it’s your turn to put that into practice.

AI Builder form-processing frequently misreads handwritten notes from certain customers, causing delays.

- Limit the use of form-processing to typed–forms only
- Improve the model with additional data
- Train branch staff to use a standard form
- Remove the form processing capability until the model performs equally

AI Builder form-processing frequently misreads handwritten notes from certain customers, causing delays.

RAI principle: Fairness, Reliability and safety

- [ ] Include human-in-the-loop validation
- [ ] Expand training data to include diverse handwriting and layouts
- [ ] Check for format variation between branches

An agent is providing incorrect product advice to numerous customers – nobody is sure who built, or approved the agent

- Add a disclaimer saying “this is not official advice”
- Track down the owner and ask them to fix it
- Establish clear ownership, review responsibilities and an update process
- Remove the agent and use a “contact us” form

An agent is providing incorrect product advice to numerous customers – nobody is sure who built, or approved the agent

RAI principle: Transparency, Accountability

- [ ] Assign clear ownership for each agent or flow within your development lifecycle
- [ ] Establish review checkpoints before publishing externally facing content
- [ ] Create a simple escalation process when content is challenged or flagged

But being proactive is always better than being reactive.

Fairness	Diversify training data; evaluate selection rates across input types
Reliability	Provide prompt design guidance; allow user validation/ editing of results
Privacy	Add DLP policies; review prompt output and access levels
Inclusiveness	Test with accessibility tools; provide fallback options and alternate inputs
Transparency	Add visual indicators (“generated by AI”); link to explanation resources
Accountability	Assign owners for review; set SLAs for update cycles and response to incidents

Embracing robust processes with Microsoft's responsible AI tools

Helps identify and assess risks before launch

Supports alignment with responsible AI principles

Use it to build a repeatable, robust AI development process

https://aka.ms/powercat/rai-template

Embracing robust processes with Microsoft's responsible AI tools

Assess fairness

Measures whether your AI model treats different groups (gender, age, or other characteristics) equally.

Identifies bias

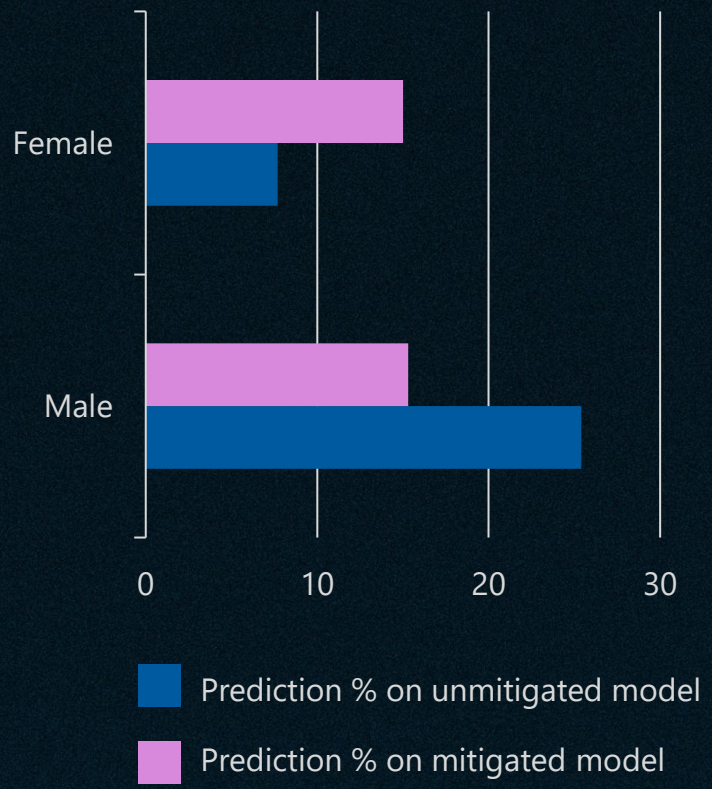
Pinpoints specific areas or ways in which bias might be occurring so you can understand the problem.

Reduce unfairness

Offers methods to adjust or retrain y model helping ensure fairer outcomes.

https://aka.ms/powercat/ai-fairlearn

Example – retraining loan application pre-approval



Unmitigated model training

- Split dataset for training and testing (60/40).
- Defined ‘sex’ as a sensitive feature.

17.7%

gap between male and female high earner predictions

Mitigated model training

- Used GridSearch to test multiple model configurations
- Used DemographicParity as a fairness constraint

0.3%

gap between male and female high earner predictions

https://aka.ms/powercat/bias-mitigation

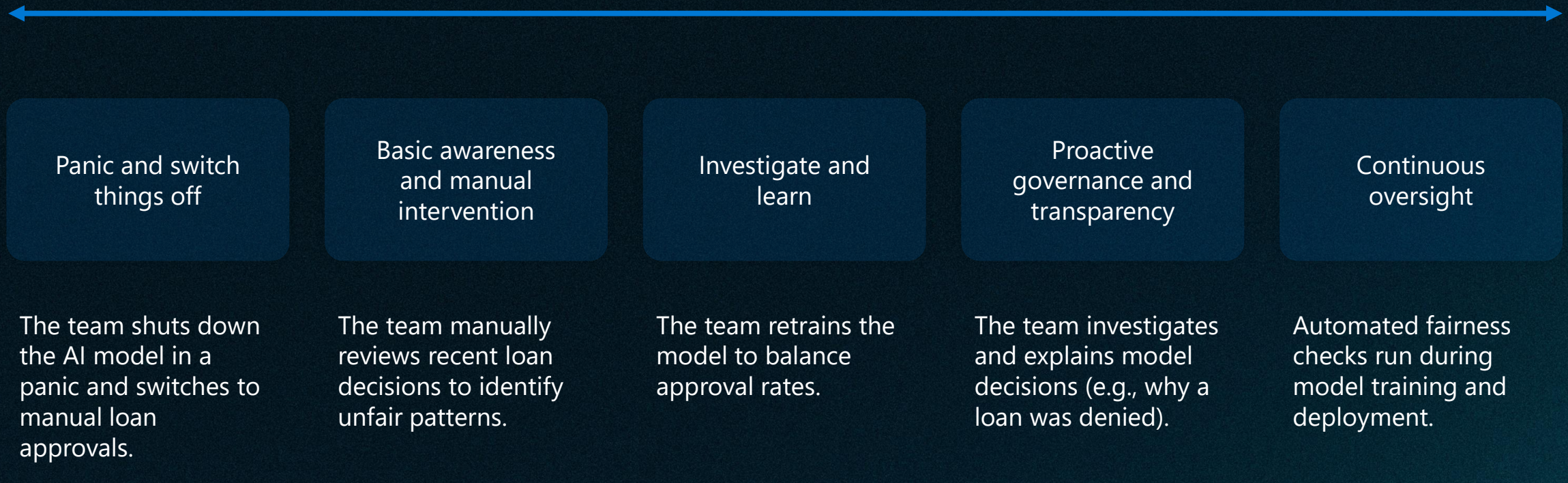
Let’s revisit our last question

Panic and switch things off	Basic awareness and manual intervention	Investigate and learn	Proactive governance and transparency	Continuous oversight
The team shuts down the AI model in a panic and switches to manual loan approvals.	The team manually reviews recent loan decisions to identify unfair patterns.	The team retrains the model to balance approval rates.	The team investigates and explains model decisions (e.g., why a loan was denied).	Automated fairness checks run during model training and deployment.

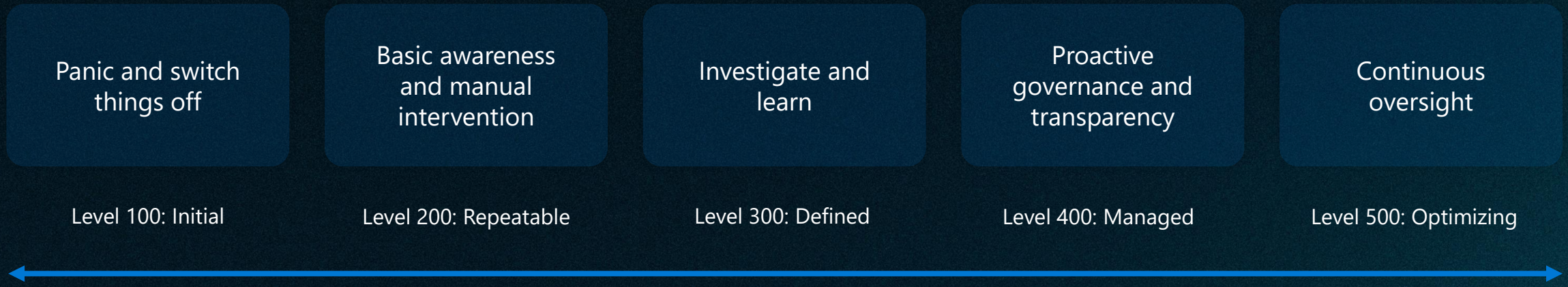
How would your organization respond?

Panic and switch things off	Basic awareness and manual intervention	Investigate and learn	Proactive governance and transparency	Continuous oversight
The team shuts down the AI model in a panic and switches to manual loan approvals.	The team manually reviews recent loan decisions to identify unfair patterns.	The team retrains the model to balance approval rates.	The team investigates and explains model decisions (e.g., why a loan was denied).	Automated fairness checks run during model training and deployment.

An indicator of capability



Also known as a maturity model

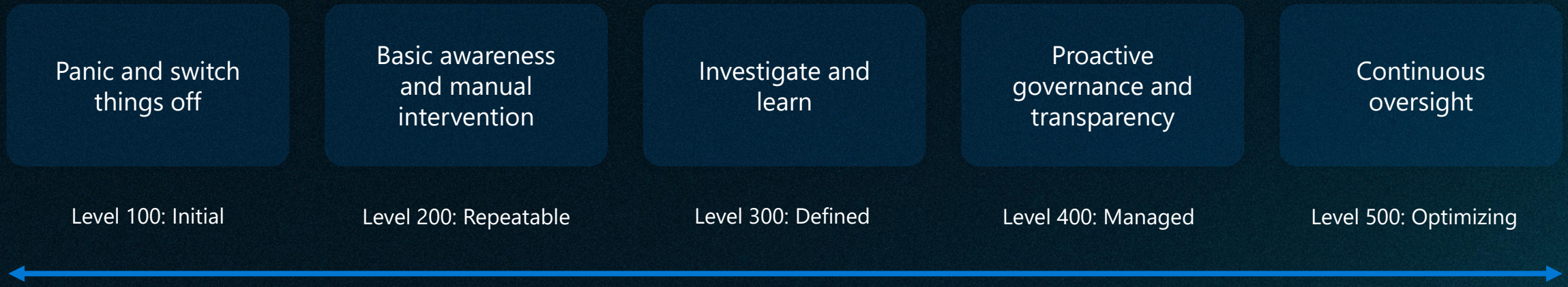


Panic and switch things off	Basic awareness and manual intervention	Investigate and learn	Proactive governance and transparency	Continuous oversight
Level 100: Initial	Level 200: Repeatable	Level 300: Defined	Level 400: Managed	Level 500: Optimizing
Process	**Process**	**Process**	**Process**	**Process**
No formal review process for AI use, ad hoc reactions to issues.	AI use cases logged and categorized, basic checklists for risk awareness introduced.	Defined AI review gates, AI impact assessments conducted before deployment.	Responsible AI principles embedded in lifecycle; incident response plans established.	AI risk management is part of organizational strategy and performance measurement.

Panic and switch things off	Basic awareness and manual intervention	Investigate and learn	Proactive governance and transparency	Continuous oversight
Level 100: Initial	Level 200: Repeatable	Level 300: Defined	Level 400: Managed	Level 500: Optimizing
Tools	**Tools**	**Tools**	**Tools**	**Process**
Power Platform and M365 Copilot used without structured oversight.	Sensitivity labels, DLP basics, CoE Starter Kit partially implemented.	AI model cards, Fairlearn, transparency indicators, Power Platform Admin Center.	SHAP, LIME for explainability, automated monitoring, integrated risk scoring.	Enterprise-wide AI inventory and dashboards; custom tooling integrated with development pipelines

Panic and switch things off	Basic awareness and manual intervention	Investigate and learn	Proactive governance and transparency	Continuous oversight
Level 100: Initial	Level 200: Repeatable	Level 300: Defined	Level 400: Managed	Level 500: Optimizing
Team structure	**Team structure**	**Team structure**	**Team structure**	**Team structure**
No clear ownership or accountability for AI risk or outcomes.	AI advocate nominated, early coordination between IT and business teams.	CoE includes RAI roles, model and prompt owners assigned, regular governance review begins.	Cross-functional RAI working group, active risk feedback loops.	Executive oversight of AI ethics, RAI responsibilities part of role description

Let’s run an exercise to see where we really are.



https://aka.ms/powercat/ai-risk-roadmap

A summary…

AI systems don't fail in obvious ways

They rarely show up with a label like "fairness issue" – you need to spot signals

Responsible AI is more than a principle

Each principle gives you a lens to diagnose problems and a playbook for how to fix them.

Don't just react - prepare

Having response plans ready for each principle means you're not starting from scratch when things go wrong.

The right tools already exist – use them

Microsoft' Responsible AI Impact Assessment template gives you a robust framework to build processes around.

You don't need to fix everything – just start fixing something

Progress comes from building awareness, asking better questions and embedding responsible AI into everyday decisions.

Before we get to Q&A, please provide your feedback

aka.ms/AIRisks/Feedback

Ready for more?

April 16th: **Evolving governance for AI**

April 30th: **Data foundations for AI**

aka.ms/PowerCAT/AiWebinars

Get started today

aka.ms/**trycopilotstudio**

Learn more

Copilot Studio website:	aka.ms/**copilotstudio**
Blog:	aka.ms/copilotstudio**blog**
Public Demo:	aka.ms/**copilotstudiodemo**
Learn Docs:	aka.ms/**copilotstudiodocs**
Community page:	aka.ms/**copilotstudiocommunity**
Copilot Studio Resources:	**aka.ms/copilotstudio/resources**

Thank you for participating!