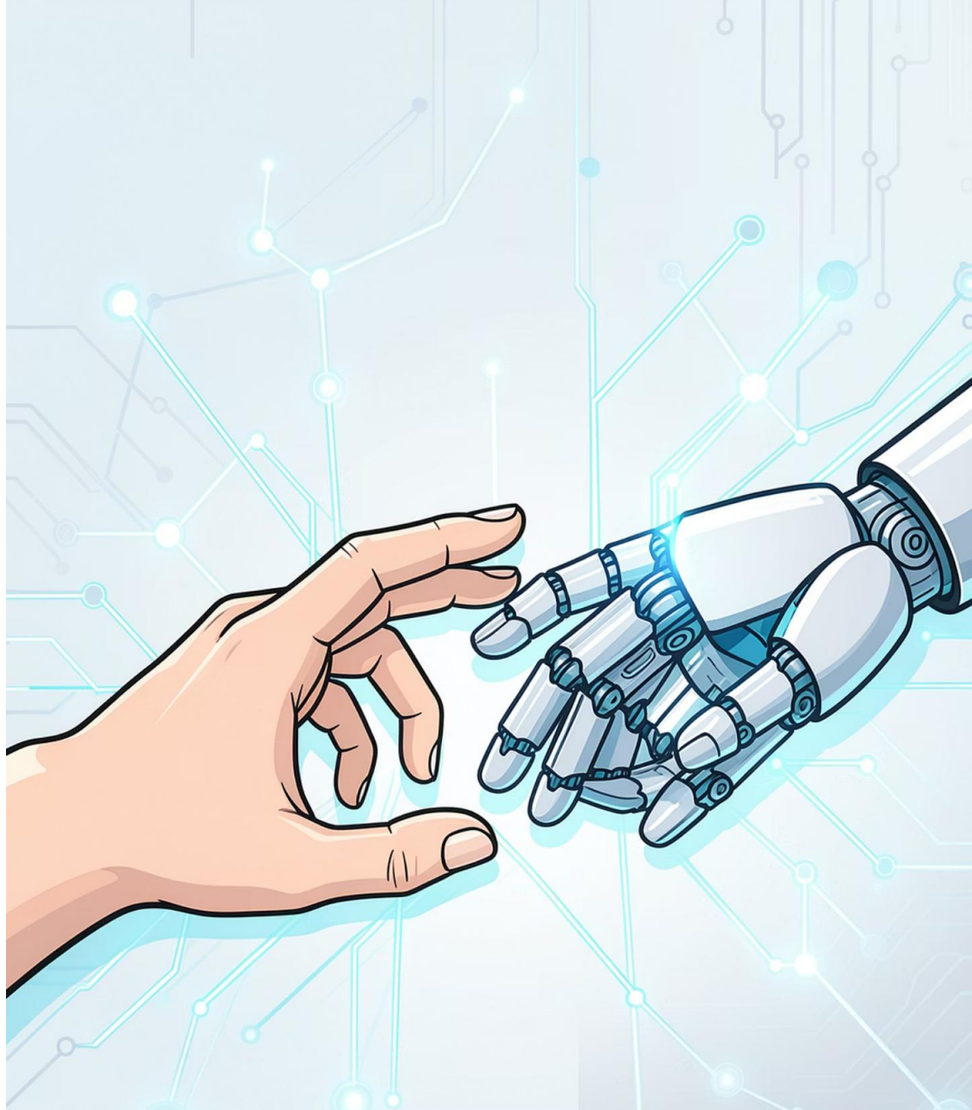
Who Is Responsible for AI Decisions?

What we need to know about responsibility, risk, and impact in the age of intelligent systems.

Susana Brás

(susana.bras@ua.pt)

universidade de aveiro | deti | IEETA | universidade de aveiro comissão de ética para a investigação clínica



An AI language model (Gamma) was used to support the creation of the presentation.
All arguments, and final wording were critically reviewed, edited, and validated by the author.

What We'll Explore Together

01

What Is Responsibility?

Defining the link between decisions, actions, and consequences

02

AI Already Decides

Recognizing invisible AI in everyday life

03

Where Does Blame Fall?

Tracing the chain of accountability

04

Key Risks

Bias, opacity, automation, and scale

05

AI in Academia

Relevance to universities and research

06

What To Do

Actionable steps for critical engagement



AI Already Makes Decisions — Even When We Don't Notice

"How many of the decisions in your day are no longer entirely your own?"

Recommendations

What you watch, read, and buy — continuously shaped by algorithms working in the background.

Filters & Rankings

Search results, job listings, and loan approvals — ordered and filtered before you ever see them.

Admin Decisions

Admissions, benefits, and institutional processes — increasingly mediated by automated systems.

Before Any Definition

Do you recognize these situations?

📺 Recommendation

A system suggested a piece of content to you. You clicked. You never questioned why it appeared — or what logic placed it in front of you.

📋 Classification

A profile was automatically assessed. The score seemed objective, authoritative, and final. You accepted it without scrutiny.

✍️ Suggestion

A text was completed by the tool. It stayed. It seemed natural — indistinguishable from your own voice.

⚠️ **If we don't know where the decision is, we also don't know who is responsible.**

The Illusion of Control

AI functions as a system of **invisible mediation** — it stands between us and our decisions without us noticing. There is a fundamental difference that is rarely discussed:

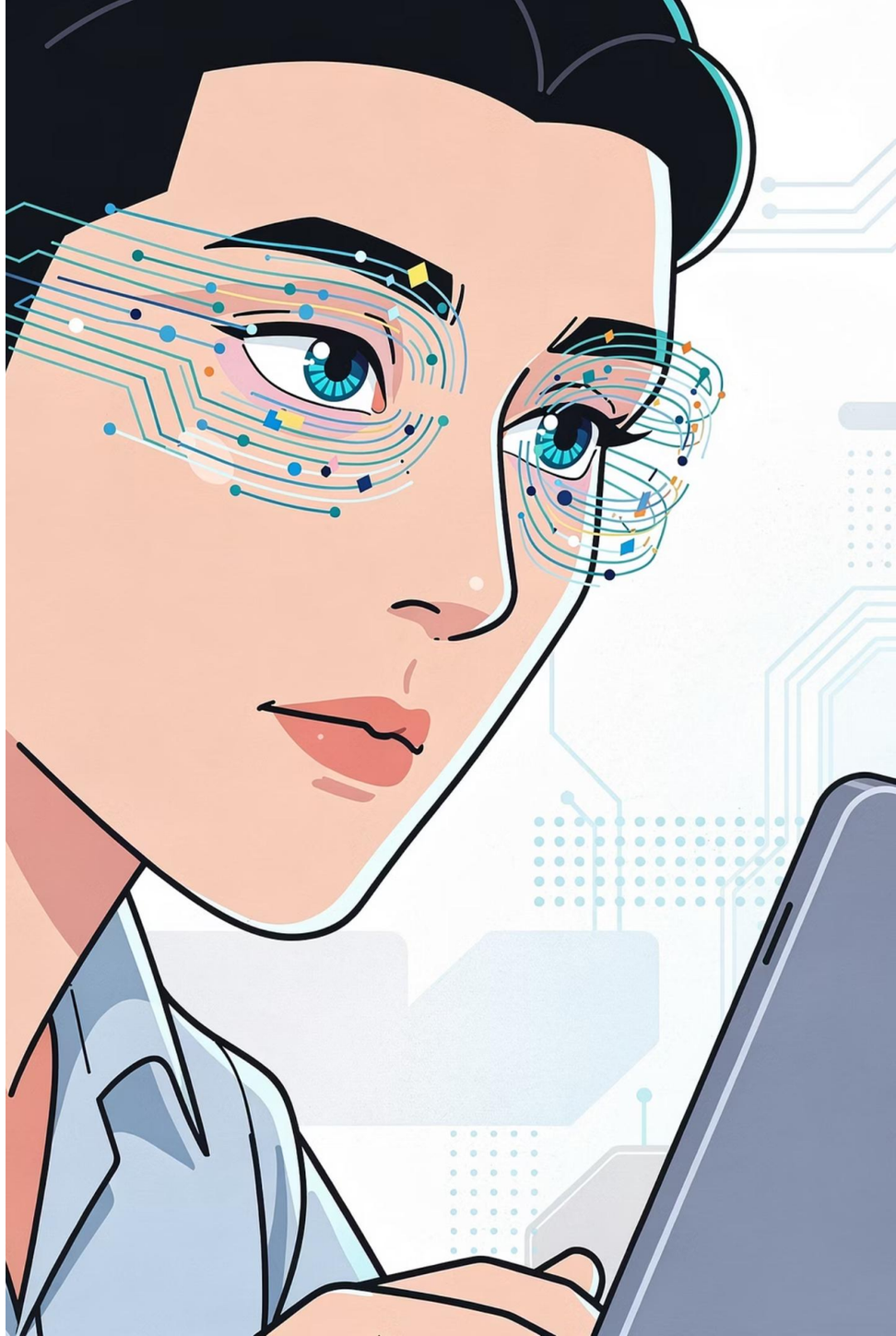
Decision Support

The system provides information. The human weighs it, questions it, and decides consciously.

Decision Substitution

The system generates an answer. The human validates it without questioning. Responsibility dissolves silently.

> The more convenient the system is, the less we question it. And that is precisely where the problem begins.





TOPIC 1

What Is Responsibility?

Responsibility means being **answerable** for the consequences of a decision or action. Technology does not remove that link — it redistributes it across a complex chain of actors, systems, and institutions.

Decision → Action → Consequence

Every system embodies choices made by someone

Technology Requires Literacy

Critical thinking is not optional — it's essential

EXERCISE 1

Mapping Your Responsibilities

Take 2 minutes to reflect on your own work. This quick exercise will anchor everything we explore today in your real professional context.

Step 1 — Identify Your Tasks

Think of **one high-responsibility task** and **one low-responsibility task** from your current project or daily work. Write both down.

Step 2 — Justify the Difference

On a separate layer, note **one or two reasons** that explain why one task carries more responsibility than the other. What makes the stakes higher?

Keep these tasks in mind — you will return to them at the end of the session.

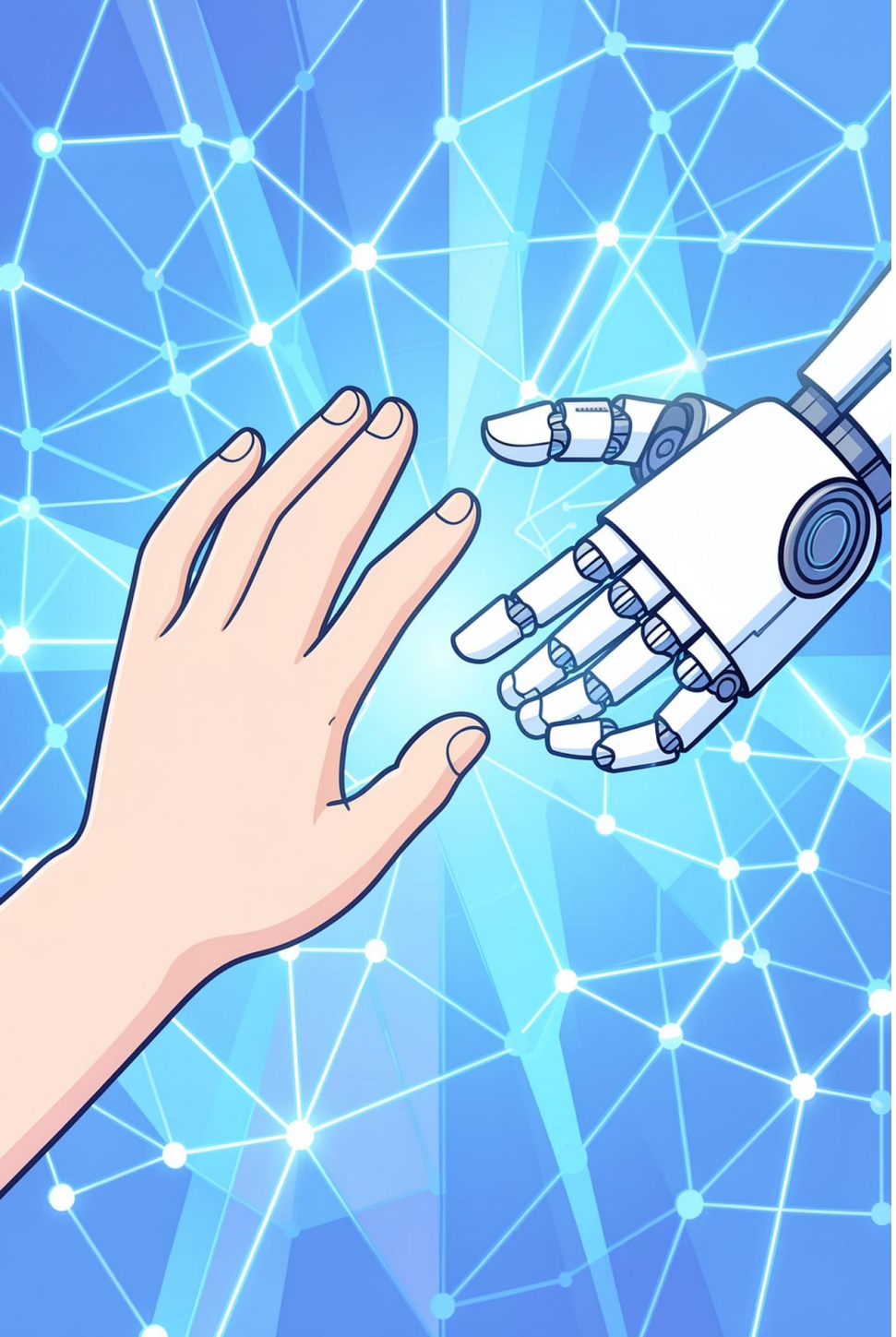


Ethics Is Not Simple — and Neither Is AI

What we call "right" and "wrong" is shaped by social and cultural contexts — and those contexts carry biases and blind spots accumulated over generations. AI systems inherit exactly those imperfections.

They learn from data generated by society itself: data that reflects historical inequities, cultural assumptions, and structural limitations. The result is not a neutral machine — **it is a mirror**. And mirrors do not correct what they reflect.

> Understanding AI ethics requires understanding the social conditions that produced the data — not just the model architecture built upon it.



AI Ethics: Building the Foundation for Responsible Innovation

As artificial intelligence reshapes industries and societies, the question is no longer *can* we build these systems — but *how* we build them, and *who* they serve. Ethics is not a constraint on progress. It is the condition for sustainable, trustworthy, and human-centered AI.



The ethical risks of AI are not primarily technical failures. They are **failures of governance, responsibility, and moral imagination.**

Developing AI Systems Is Not Just Engineering

It is participation in social decision-making — one that determines who benefits, who is harmed, and whose voice is heard in the design process.

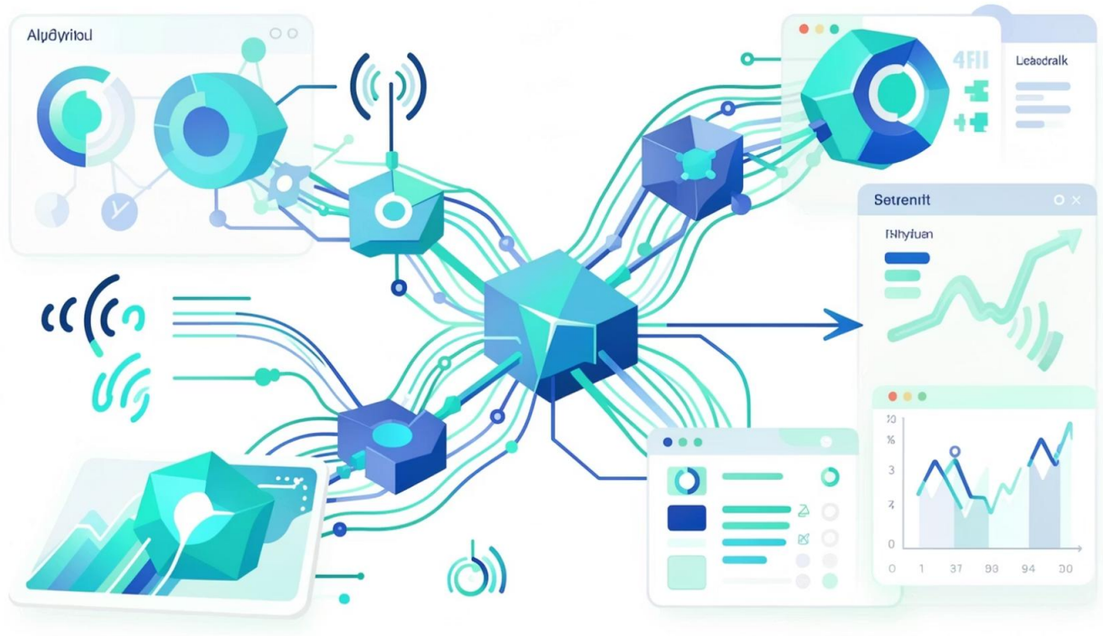
Responsibility Cannot Be Outsourced to the Algorithm

Delegating moral judgment to a system is itself a moral choice — and one that carries full accountability for its consequences.

AI Shapes the Future Distribution of Power

AI systems do not simply reflect society; they actively reconfigure who has access, influence, and opportunity in the years ahead.

Algorithms Optimize — They Don't Judge



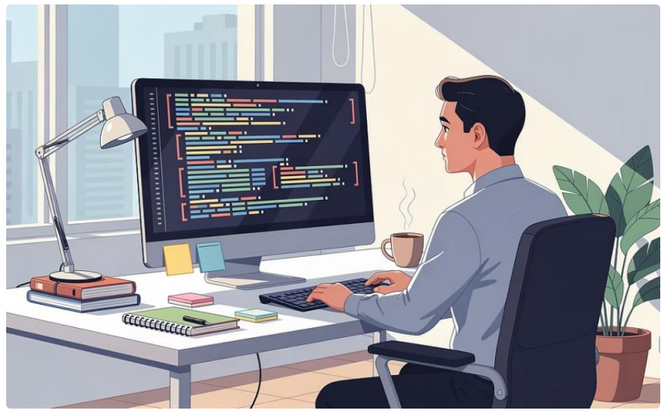
AI systems are designed to **minimize error and maximize performance** — not to uphold justice or fairness. This is a critical distinction that practitioners and policymakers must internalize.

If a biased solution produces a lower statistical error, the algorithm will reproduce it without hesitation. It is doing exactly what it was designed to do. The problem is not merely technological. It is **structural** — and it demands both technical understanding and deep contextual knowledge to address responsibly.

Optimization for accuracy is not the same as optimization for fairness. These goals can — and frequently do — conflict.

Two Dimensions of Critical Capacity

To interpret, audit, and challenge algorithmic decisions effectively, we need both technical literacy and domain expertise. Without either dimension, critical capacity is incomplete — and accountability remains out of reach.



Technical Understanding

A working grasp of what AI is, how it operates, how models are trained, and where their limitations lie — even for non-specialists. This enables meaningful scrutiny of automated outputs.



Problem Context

A deep understanding of the specific problem being solved — its social, ethical, and human dimensions. Without this, even technically sound systems can produce unjust outcomes.

Together, these two dimensions form the foundation for responsible AI governance — and neither can substitute for the other.

EXERCISE 2

When AI Gets It Wrong — Who Is Responsible?

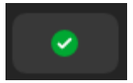
Imagine this scenario: an AI platform is embedded in your project management system. It works seamlessly, and your team relies on its suggestions daily. Then, one day, you realize the platform is biased — its recommendations are sometimes out-of-context or misleading.

Who is responsible? Consider each candidate below, then vote for your answer.

1

The Designer

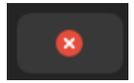
Who conceptualized and designed the solution



2

The Developer

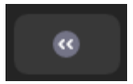
Who built and deployed the solution



3

The User

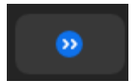
Who used the solution without questioning it



4

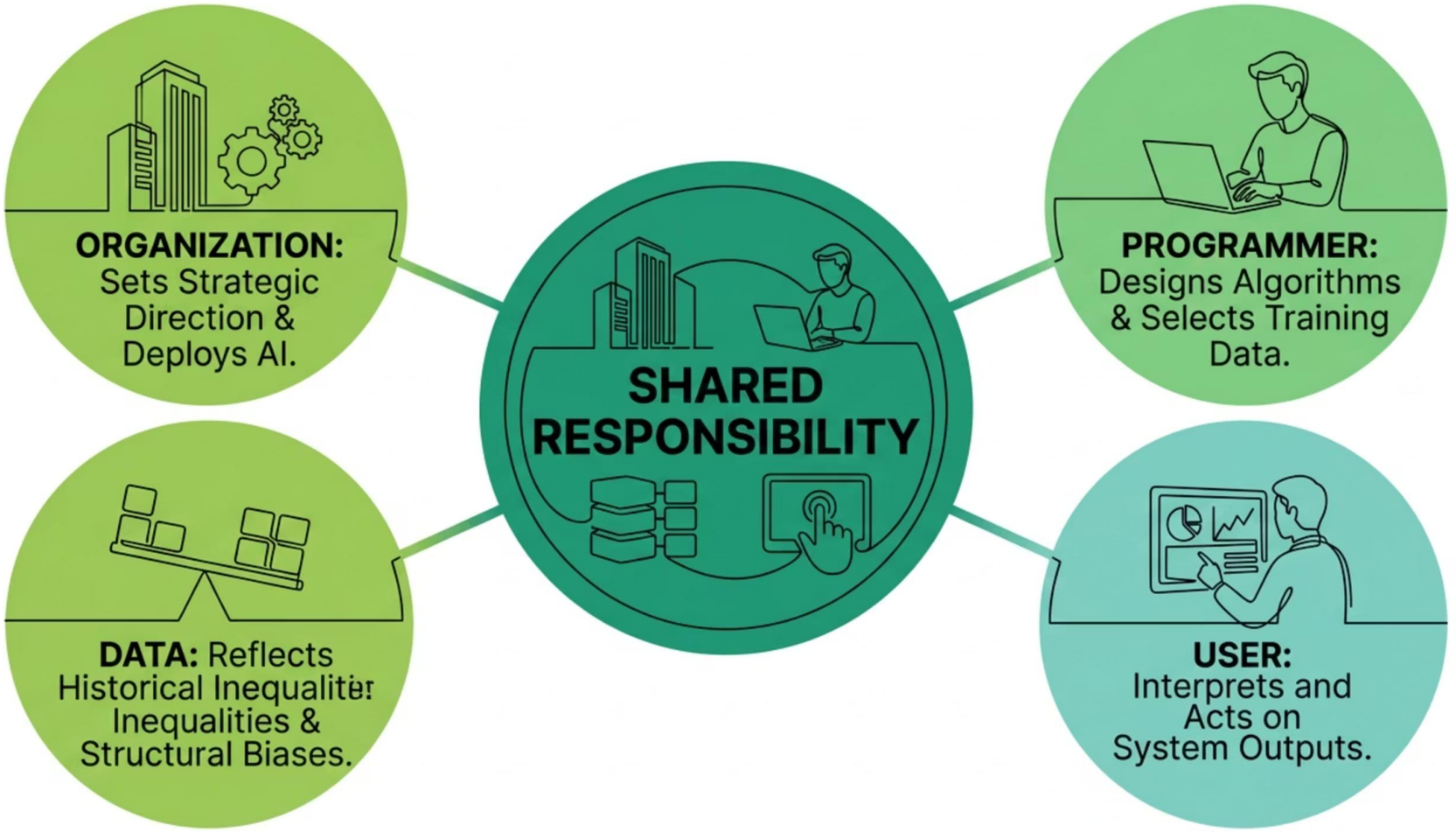
All of them

A distributed responsability



TOPIC 2

When Something Goes Wrong — Where Does Responsibility Lie?



Critical insight: Responsibility ≠ simple individual blame. The system itself — and the conditions that produced it — can be the primary source of harm.

When Something Goes Wrong

Imagine: an automated system denies a loan, rejects an application, or assigns a lower score. It happened. The harm is real. And now — who is responsible?

1

Who Develops

Defines the algorithms, chooses the training data, establishes the optimization objectives — and determines what "success" means for the system.

2

Who Implements

Adopts the system, defines the usage context, and determines the scope of its recommendations within a specific organizational environment.

3

Who Uses

Follows or ignores the system's suggestions, with greater or lesser critical judgment — often under institutional pressure to defer to automated outputs.

4

Who Trusts

Accepts results as legitimate, transferring authority to the system without scrutiny — creating conditions where errors go unchallenged.

⚠ **AI does not eliminate responsibility — it redistributes it in a barely visible way.** Resist the temptation to find a single culprit.

Responsibility Is Shared Across the Pipeline

The debate on AI responsibility has driven the creation of methodologies to integrate ethical practices throughout the **entire development and deployment pipeline**. Distributed accountability ensures ethical considerations are embedded from conception to utilization — not bolted on as an afterthought.



Those Who Conceptualize

Entities and teams that define the problem, establish ethical objectives, and determine what outcomes the system should — and should not — produce.



Those Who Develop

Engineers, data scientists, and technical decision-makers who design, train, validate, and deploy models — making consequential choices at every step.



Those Who Use

Organizations, professionals, and citizens who interact with systems and interpret results — bearing responsibility for the judgment they apply (or withhold).

TOPIC 3 — CORE RISKS

The Four Fundamental Risks of AI

Bias

Data reflects historical inequalities. AI doesn't merely mirror the world — it can **amplify** its injustices at unprecedented scale.

Opacity

Many AI systems are **"black boxes"** — decisions are difficult to explain, audit, or meaningfully challenge by those affected.

Uncritical Automation

Humans naturally trust systems. We often accept automated outputs **without questioning** them — surrendering judgment to the appearance of objectivity.

Scale

A small error, deployed at scale, becomes a **systemic failure** affecting thousands or millions of people — often before anyone notices.

Why Does AI Fail — Even When It "Works"?

You don't need to understand how it works technically to understand where it fails conceptually. Three ideas that cannot be ignored by anyone engaged in AI governance:

1 Bias Is Not an Exception — It's Structural

Data reflects the real world — with all its historical inequalities and patterns of exclusion. A system trained on biased data learns and reproduces those biases, even without intent or malice from its designers.

2 Opacity Limits Accountability

When we cannot explain how a decision was made, it becomes extremely difficult to hold anyone responsible. The black box protects the system — not the people affected by it.

3 Automation Creates Excessive Confidence

Consistent research shows that humans tend not to question automated systems, even when intuition would suggest otherwise. The appearance of objectivity is, itself, a powerful form of influence.

⊗ **A system can be technically correct and socially wrong.** This distinction is at the core of every serious ethical discussion about AI.

Every AI System Embeds Human Decisions



Incomplete Mirrors

Our datasets, like our societies, often reflect only part of the whole. What problem is worth solving? What data is collected? Whose experiences are represented? What counts as success — and whose failures are deemed acceptable?



Data Reflects History, Not Truth

Datasets encode the historical record of social inequalities — in hiring, policing, credit scoring, and healthcare access. Training an AI system on biased data doesn't create objectivity. It **automates injustice**.



Scale Amplifies Everything

A correlation shaped by structural discrimination becomes a decision rule applied to millions. Machine learning amplifies patterns — it does not merely replicate them. The result: **systemic harm at speed**.

Representation is not a one-time fix — it is a continuous, ethical commitment requiring ongoing awareness, documentation, and intervention.



Questioning Technology Is Not Rejecting It

Questioning does not mean limiting or discrediting technology — it means **deepening knowledge, assessing implications, and ensuring that AI adds value safely and fairly**.

The future of AI depends on the trust society places in these systems. That acceptance will only be sustainable if technology can be audited not just by specialists, but by everyday users, institutional leaders, and affected communities.

This requires **universal design**: systems built to minimize risk, enable scrutiny, and operate with sufficient transparency for meaningful human oversight.

WHY IT MATTERS

These Risks Are Not Abstract

They affect real people in real decisions — **hiring, lending, healthcare, education, and justice**. Each of these domains involves consequential outcomes that shape life opportunities and determine who flourishes.

Individual Harm

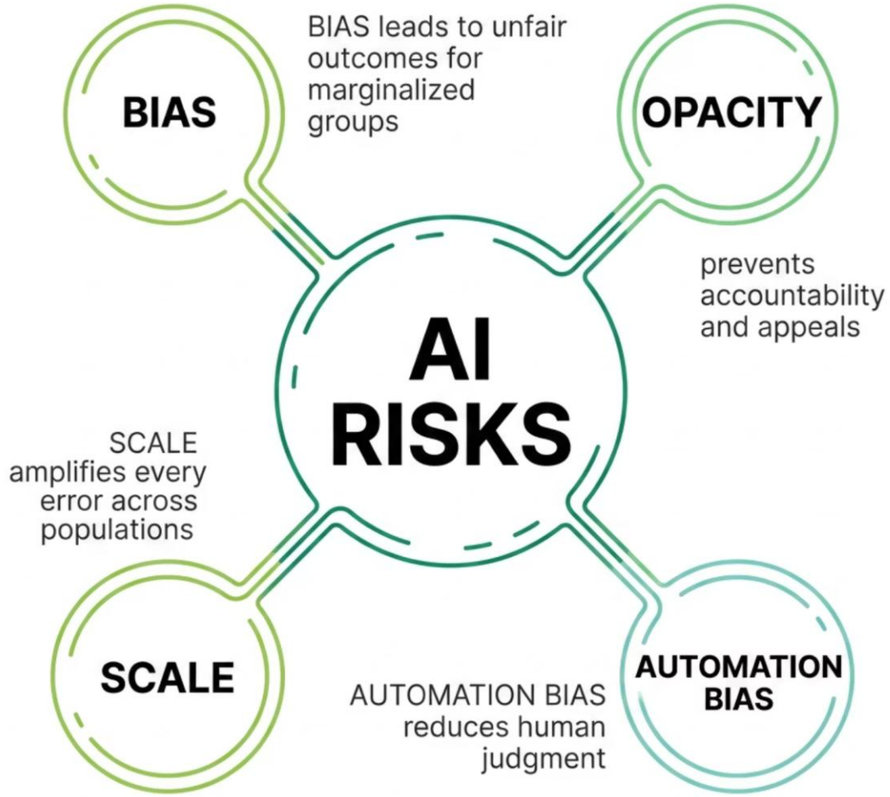
A single biased decision — a rejected application, a misclassified risk score — can close doors that are difficult or impossible to reopen.

Systemic Harm

When bias and opacity combine with scale, individual errors become patterns — and patterns become structural conditions that reinforce existing inequalities.

Erosion of Trust

Once communities experience systematic harm from AI systems, rebuilding trust in institutions that deploy them becomes an enormous challenge.





Higher Education's Decisive Role

Achieving a truly "universally designed" AI ecosystem—one that minimizes risks, enables scrutiny, and operates with transparency—hinges on the capabilities nurtured within our universities and research institutions.

Shaping Future Leaders

Equipping the next generation of AI developers, ethicists, and policymakers with critical skills and ethical frameworks.

Fostering Interdisciplinary Research

Bridging technical advancements and ethical implications through collaborative research across diverse fields.

Promoting Public Understanding

Empowering citizens and professionals to understand, question, and engage with AI technologies effectively.



TOPIC 4

And at the University?

AI is already embedded in academic life. The question is no longer whether it is present — but whether we engage with it **critically or passively**.

Automated Assessment

Grading tools and plagiarism detectors make consequential decisions about student work — often without transparent criteria or meaningful appeals processes.

AI in Research

Tools shape what questions get asked, which sources surface, and how findings are synthesized — influencing the very production of knowledge.

Student Use of GenAI

Generative tools are reshaping learning, writing, and thinking — often faster than institutions can develop coherent policies to guide their use.

> "If a system makes a mistake in an academic decision, who is accountable?"

The University Is Not Neutral

AI is already entering academic institutions — often without debate, without clear policy, without awareness of the risks. Think about your own contexts:

Assessment

Plagiarism detection tools, automated grading, performance analysis.

Research

Text generation, literature synthesis, grant application screening.

Student Tools

ChatGPT, Copilot, automatic translators — used daily, without ethical guidance.

Administrative Decisions

Scholarships, admissions, resource allocation — decision-support systems with real impact.

⚠ If an academic decision is influenced by AI — who takes responsibility for the error?



EXERCISE 3

Scenario: Merit Scholarships and AI

A university implements an AI system to identify merit scholarship applications with the highest likelihood of success. The system does not decide — it generates a **ranking and a recommendation** for the evaluation committee. After a few months, a pattern emerges.

1

Implementation

The system is adopted to increase efficiency and reduce workload. Few questioned the composition of the training data or the definition of "merit" embedded in the model.

2

Regular Use

The committee follows the recommendations. The ranking seems objective, consistent, and authoritative. Confidence in the system grows — and critical scrutiny declines.

3

Pattern Discovered

Months later: students from disadvantaged socioeconomic backgrounds are systematically ranked lower — regardless of academic achievement or potential.

4

The Question

Who is responsible? The developers? The committee? The institution?

This scenario is not hypothetical. Variants of it are playing out in institutions around the world — often without anyone claiming accountability.

The Questions

1

Who is responsible?

Who developed it?
The university?
The committee that used the recommendation?
Who approved the purchase?
Nobody?
Everyone partially?

The goal is NOT consensus

The Questions

2

If the system only "helped," why did it influence so much?

Uncritical automation.

Algorithmic authority.

Excessive trust in the system.

A system does not need to decide to shape decisions.

The Questions

3

Would you use the system if:

- ✓ The system reduces analysis time by 70%
- ✓ but it may introduce subtle biases.

The real decision is between:

- ➢ Efficiency vs. fairness
- ➢ Convenience vs. risk

The problem is rarely just technical.

The Questions

1

Who is responsible?

Who developed it?
The university?
The committee that used the recommendation?
Who approved the purchase?
Nobody?
Everyone partially?

2

If the system only "helped," why did it influence so much?

Uncritical automation.
Algorithmic authority.
Excessive trust in the system.

3

Would you use the system if:

- ✓ The system reduces analysis time by 70%
- ✓ but it may introduce subtle biases.

The real decision is between:

- ➢ Efficiency vs. fairness
- ➢ Convenience vs. risk

The goal is NOT consensus

A system does not need to decide to shape decisions.

The problem is rarely just technical.



Responsibility Is Shared — Not Single

In most real-world scenarios, **there is no single answer** — responsibility is distributed across all entities involved.

Is there a single responsibility?

Rarely. Accountability is layered across designers, developers, deployers, and users.

Is there a unique response?

No. Context shapes the appropriate response — the same AI failure can demand different actions in different settings.

What should happen?

Not abolish — but **use with care**. Applying AI without scrutiny is the real risk.

> ⚠ In front of different situations, should we always hold the same posture?

Before Using AI: The Moment That Almost Never Exists

Instead of a list of “best practices,” we propose a simple mental framework — four questions to ask *before* delegating any decision to an automated system. It’s not about using AI better. It’s about deciding whether it should be used at all.

1

What am I delegating?

Identify precisely what stops being done by me.

2

What could go wrong?

Even if unlikely — what are the failure scenarios and their impacts?

3

Who is affected?

Map the people who will be impacted by the decision mediated by the system.

4

Can I justify it without AI?

If I can’t explain the decision without the system, maybe I shouldn’t make it with it.

EXERCISE 4

Back to Your Tasks — Does AI Fit Both?

Remember the high- and low-responsibility tasks you identified at the start? Now ask yourself:

Do you use AI in both?

Is the strategy the same?

What determines the care?

EXERCISE 4

Back to Your Tasks — Does AI Fit Both?

Remember the high- and low-responsibility tasks you identified at the start? Now ask yourself:

Do you use AI in both?

Most people do — or could. But the real question is whether you should approach both tasks the same way.

Is the strategy the same?

Probably not. The higher the stakes, the more care, validation, and human judgment must accompany AI use.

What determines the care?

The **degree of compromise and responsibility** attached to your task. This is the real driver of how you should use AI.

Different tasks demand different care, different scrutiny, and different tools. Understanding your task deeply — what it means and what consequences it carries — determines the way you will responsibly use AI.

TOPIC 5

What Do We Realistically Need to Do?

Not a long list of rules — but a shift in stance. Critical engagement with AI is now a core professional and civic competence.

Question Before You Use

Ask: What data trained this? What does it optimize for? Who benefits?

Demand Human Oversight

Automated decisions, especially high-stakes ones, require a human in the loop

Champion Diversity

Diverse teams and data reduce blind spots and embedded bias

Build AI Literacy

Understanding how AI works is no longer optional — it's transversal



Your AI Toolkit: Beyond the Prompt

AI outputs are not magic. They are products of your input, your judgment, and your choices.

Ask yourself: "Would I be comfortable explaining this AI-generated outcome to my supervisor, colleagues, or students?"

Before using an AI tool, ask: "Does this align with our institutional mission? Am I protecting sensitive data?"

Your Role Is Not Neutral

Whether you are a **user**, a **decision-maker**, or a **creator** — you are already part of the AI system. The question is whether you participate consciously.

Recognize

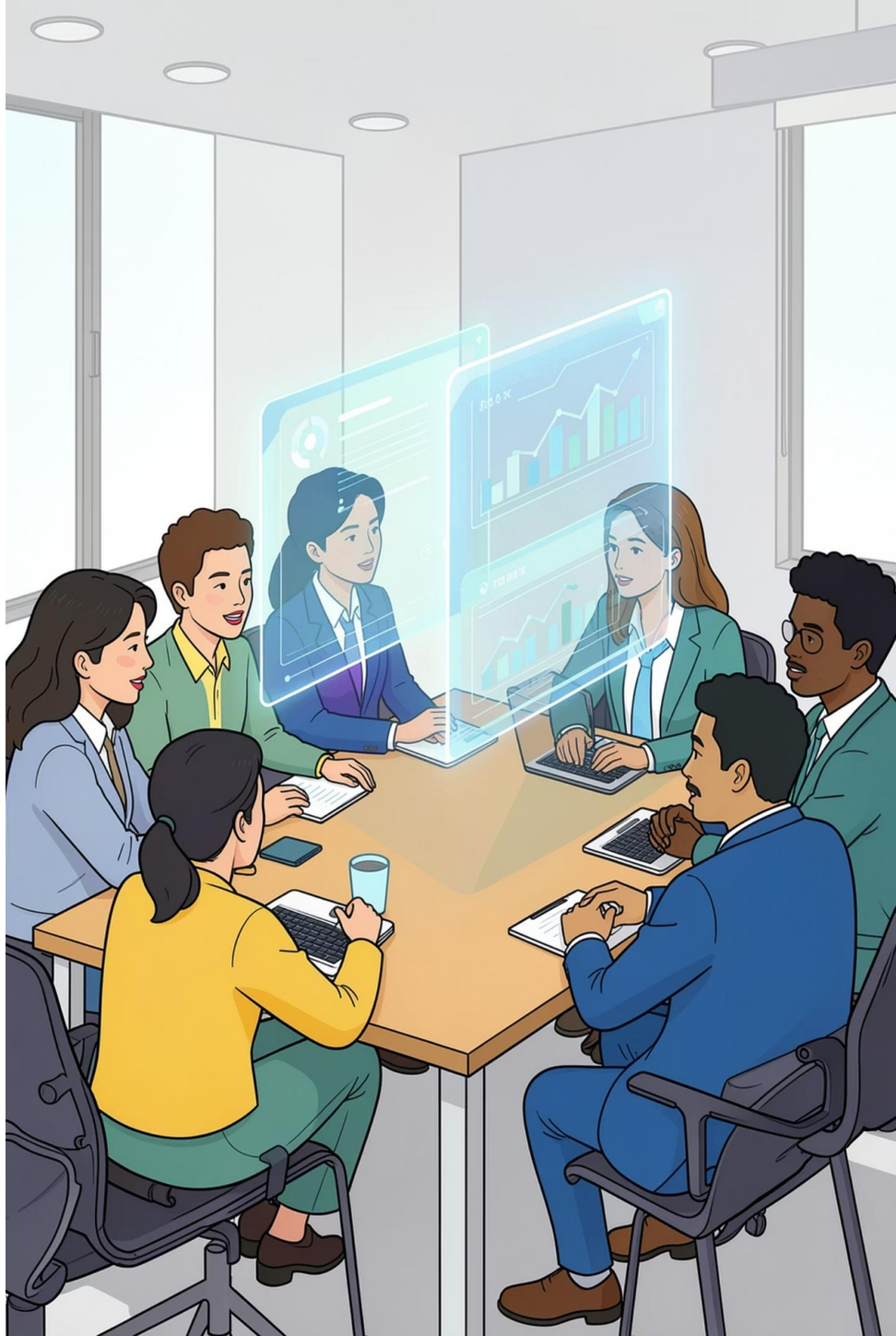
See where AI is making decisions around you

Question

Challenge opacity, bias, and uncritical trust

Act

Demand accountability — in your institution and beyond





When Not to Use Generative AI

Some categories of information should **never** be entered into a generative AI platform. This is a firm boundary, not a suggestion.

Confidential Information

Internal strategies, contracts, client data, or any content marked as restricted within your organization.

Personal Data

Names, identifiers, contact details, or any information that could be traced back to a real individual.

Sensitive Data

Financial records, health information, legal matters, or any data with regulatory or reputational implications.

These models work with data — *your* data. Every interaction feeds the platform. Confidential information should never be deployed into a machine that will use and potentially expose it.

When Extra Care Is Required

If your task involves sensitive content but still benefits from AI assistance, proceed only with strict precautions in place.

Sanitize Inputs

Remove confidential and personal content before sending



Use Approved Platforms

Only use institutionally approved tools with verified policies

Institutional platforms provide a privacy policy and data governance layer that public tools do not. Even so, the discipline of removing sensitive content before prompting is non-negotiable — the platform's policy does not protect information you actively share with it.

High-Stakes Tasks — Use Critically

Ask These Questions First

Before you open any AI tool, assess the stakes of the task:

- Does my task **directly affect someone**? → Use critically
- Does my task **influence decision-making**? → Use critically

The higher the human impact, the more rigorous your validation must be.

The Pre-AI Planning Protocol

Before running any prompt, complete these steps:

1. Plan the task's steps end-to-end
2. Schematize the workflow visually
3. Do a preliminary analysis where applicable
4. Define what the expected result looks like

Then use GenAI — and compare its output against your own intuition.

⚠ **Why plan first?** If you look at the AI response before forming your own view, you will likely accept it without question. Generative AI is extremely convincing — which is precisely what makes it dangerous when used without prior reflection.

A Living Checklist

This is not a finite list. Adapt these practices to your team's context and the specific nature of your work.

1

Prepare Prompts Carefully

Clear, simple, detailed about the question, and explicit about the result you expect. A good prompt is a design artifact.

2

Save All Outputs

Preserve the AI-generated content as part of your project record, including the version and platform used.

3

Register AI Use in Deliverables

Clearly note where and how AI tools contributed to any final output — for transparency and traceability.

4

Validate Every Output

Cross-check AI results against your own expertise, available data, and the expected outcome you defined in advance.

Critical note: Worse than not using AI — or not documenting it — is having *no clear strategy at all*. Reactive, unplanned AI use creates invisible risk.

Build Your Team's AI Strategy Together

If you work in a team, you have both the opportunity and the responsibility to define a shared approach to AI use. Don't leave it to individual habit.

Start the conversation

Open a team debate: *Where is our red line? How far can we go, and about what topics?*

Co-create an AI Use Guide

Collaboratively define: what AI means for your team, which tasks AI can support, which tools are appropriate, and how to document use consistently.

Why it matters

Thinking together about AI forces clarity. It turns vague habits into deliberate strategy — aligning everyone on risk, quality, and purpose.

> Using AI **with strategy** maximizes results and minimizes risks. The goal is not to restrict AI — it is to make your team smarter about how you use it.

AI Literacy: The Foundation of a Fair Ecosystem

True AI literacy is multidimensional — it must equip citizens and professionals to engage with technology critically and responsibly.

🔬 Technical

Understanding how AI systems work, their inputs, outputs, and limitations.

⚖️ Ethical

Recognizing bias, fairness, and the moral dimensions of algorithmic decisions.

🌍 Social

Understanding how AI affects communities, economies, and social structures.

💡 Philosophical

Asking deeper questions about autonomy, agency, and the role of technology in human life.

The Hardest Part Isn’t Understanding How It Works

It’s deciding when to trust.

The discussion about ethical AI doesn’t end with regulation, doesn’t end with technical auditing, and doesn’t end with training. It ends—or begins—each time a person consciously decides how much to delegate, to whom, and whether they are willing to accept the consequences when the system fails.

When to Trust

Trust in a system should be proportional to our ability to question it and the impact of its decisions.

How Much to Trust

Total trust is the abandonment of human judgment. AI is a tool—not an authority.

Who Takes Responsibility

Someone has to be accountable. If no one is, the system should not have been used that way.

> AI is not the problem. The lack of critical thinking about AI is.

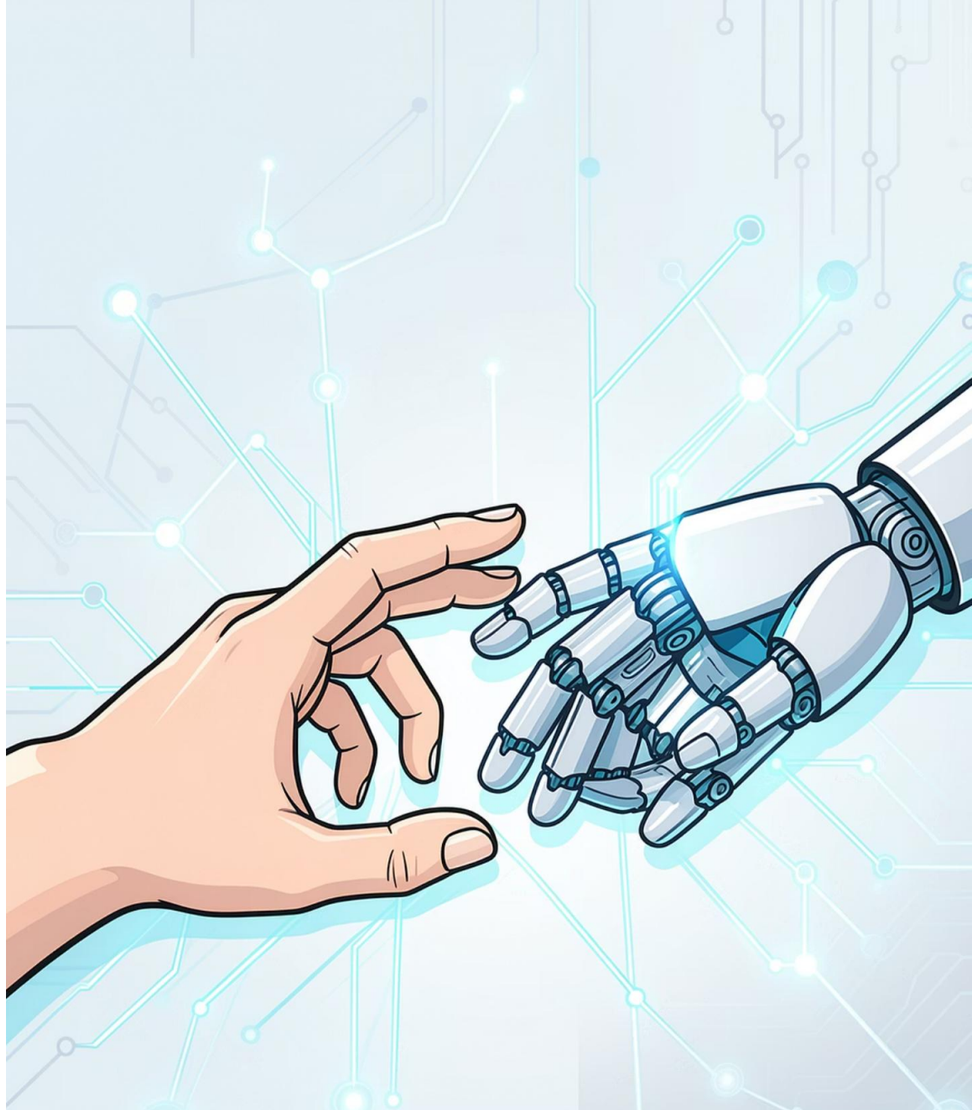
Who Is Responsible for AI Decisions?

What we need to know about responsibility, risk, and impact in the age of intelligent systems.

Susana Brás

(susana.bras@ua.pt)

universidade de aveiro | deti | IEETA | universidade de aveiro comissão de ética para a investigação clínica



An AI language model (Gamma) was used to support the creation of the presentation.
All arguments, and final wording were critically reviewed, edited, and validated by the author.