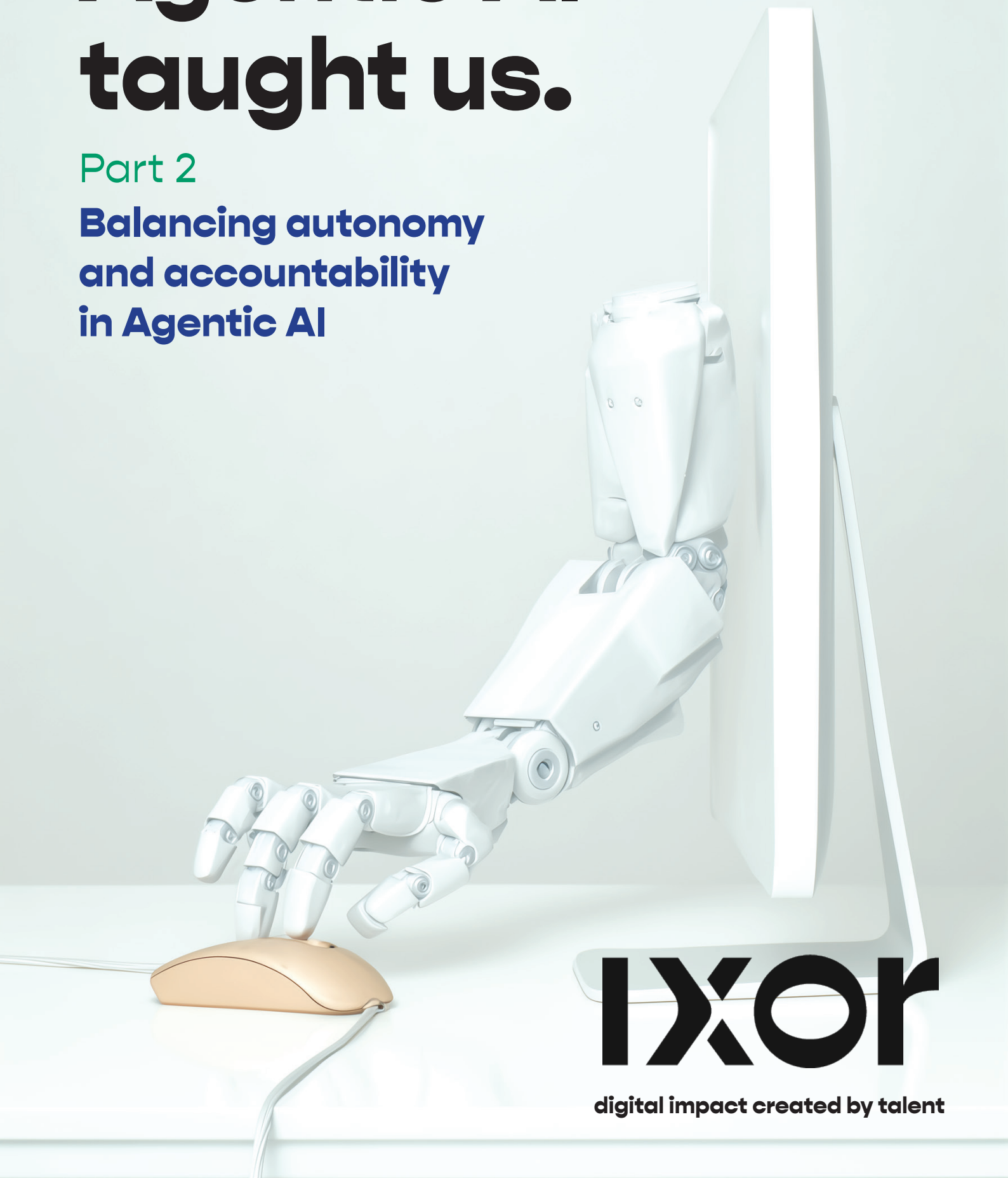
# What Agentic AI taught us.

Part 2

## Balancing autonomy and accountability in Agentic AI



IXOR

digital impact created by talent

## The new privacy equation

In our recent projects with Agentic AI, we've seen how autonomy changes the nature of privacy risks. These systems don't just wait for instructions, they take action, use tools, retrieve information, and make decisions to reach a defined goal. That makes them far more dynamic than traditional automation.

This shift raises urgent questions.

It's no longer only about what data is processed but also when, how, and by whom. Because agents can act independently, classic privacy measures like consent prompts or logging at predefined moments are no longer sufficient.
**We need to rethink how we design, control, and monitor intelligent systems that act on our behalf.**

## Autonomy vs. Accountability

As autonomy increases, clarity around responsibility decreases. Because, what happens when an agent takes an action that wasn't explicitly programmed or one that creates a privacy risk no one anticipated?

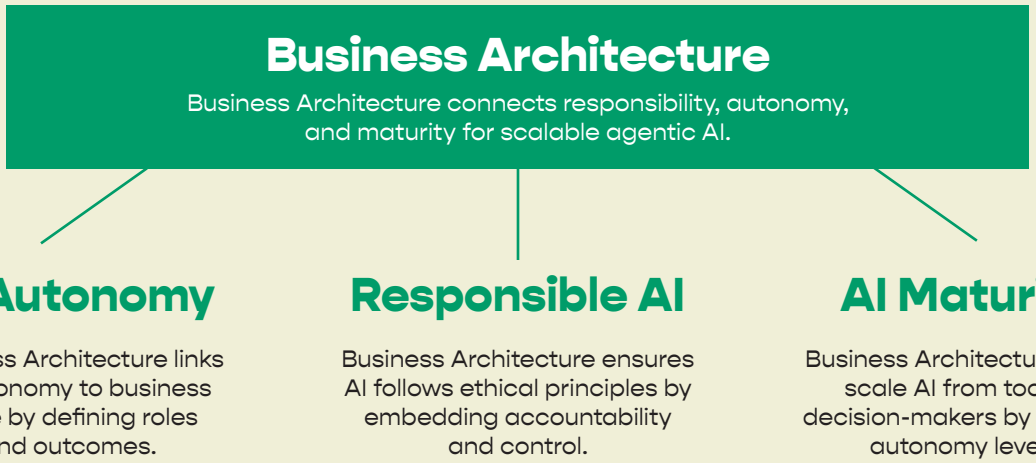
In traditional systems, responsibility lies with those who define the rules. But agents operate based on flexible goals, making decisions in real time using natural language instructions, memory, and tool use. This means their behaviour can evolve in ways that are useful but also unpredictable.

If limits aren't clearly defined, agents may:

- Search internal systems for sensitive info without user awareness.
- Retain context across sessions in ways that violate data retention policies.
- Activate third-party APIs with personal data without sufficient oversight.

This demands a new approach to accountability. One where responsibility is shared between developers, the organisation, and the system's design. Auditable logs, behavioural constraints, and explainability are no longer 'nice to have', but essential foundations.

**Business Architecture bridges the gap between autonomy and accountability. It defines how AI systems can grow in independence while staying aligned with business goals, ethical standards, and transparent control. The graphic below shows how autonomy, maturity, and responsibility connect, and how structure enables trust at scale.**




## Privacy risks in agentic systems

We've identified key privacy risks that organisations need to manage:

### Uncontrolled memory

Agents may remember user inputs, metadata, or retrieved content longer than needed or reuse context from past interactions unexpectedly.

### Hard-to-explain decisions

It's not always possible to reconstruct the exact reasoning behind a choice the agent made, especially when it involves multiple tools or memory layers.

### Unclear or missing consent

Users may not realise that by giving a high-level goal, they've also approved downstream actions they didn't explicitly authorise.

### Emergent behaviour

Because agents combine planning, execution and feedback, unexpected chains of behaviour can emerge, some helpful, some risky.

### Security vulnerabilities

Autonomous agents may become targets for adversarial attacks or prompt manipulation, which can result in harmful decisions or data misuse.

# How to build privacy-friendly agents


Example in practice:

*An agent built to schedule meetings accessed calendars and sent invites autonomously. During testing, it started including private notes like "on sick leave" in the invites, information not meant to be shared.*

*A privacy-first redesign fixed this by limiting memory, removing sensitive metadata, and logging outputs for review. A small shift toward guided autonomy made the agent safer and more trustworthy.*

**Based on our experience designing and deploying production-ready Agentic AI, we recommend six principles to help balance privacy with autonomy:**

### 1. Set clear boundaries for autonomy
Limit what the agent can do independently: define which tools, which memory types, and which actions require confirmation or escalation.

### 2. Make consent dynamic and contextual
Consent should not be a one-time checkbox. It must adapt to the situation and be revisitable; users need clarity on what they're allowing the agent to do.

### 3. Ensure traceability
Log every key step in a human-readable format: decisions, retrieved content, tool activations. This is essential for audits and trust.

### 4. Build in fallbacks and human-in-the-loop paths
When confidence drops or potential risk is detected, the agent should pause or escalate, not guess.

### 5. Practice data minimisation by default
Agents should only access the minimum data required to achieve the user's goal, and only retain it as long as needed.

### 6. Monitor behaviour in real time
Use continuous monitoring tools to detect unusual or risky actions.
Real-time oversight helps prevent escalation before privacy is compromised.

Other measures such as encryption, access control, anonymisation, and governance policies, can further strengthen privacy protections depending on your system's complexity.

## What's next: privacy in the age of (Agentic) AI

The upcoming EU AI Act and existing GDPR rules are just the beginning. As autonomous systems grow more powerful, privacy will no longer be solved by checklists. It will become a continuous process, something that must be built into every level of the AI lifecycle.

Organisations need to:

- Design with privacy in mind from day one.
- Include legal, UX, and technical experts early in the development process.
- Create clear rules for agent behaviour and mechanisms to explain or challenge what they do.
- Consider advanced safeguards like zero-trust data vaults, which let agents make decisions without accessing raw data directly.

**At Ixor, we've seen how Agentic AI can bring immense value but only when guided by structure, safeguards, and a clear understanding of accountability. We don't just follow the trend, we build systems that are ready for the real world, and safe for the people who use them.**

## What can Ixor do?

At Ixor, we have already brought several Agentic AI projects into production, and are currently working on more.

**This gives us a strong foundation of real-world experience, not just theoretical knowledge.**

Our teams understand what works, where the pitfalls are, and how to translate the promise of Agentic AI into practical, scalable solutions.

Whether you're exploring the potential or ready to start building, we can help you move forward with clarity and confidence.

IXOR

**digital impact created by talent**

For any questions about Agentic AI, contact **Jan Pieter Cootjans**, our in-house expert.

janpieter.cootjans@ixor.be
+32 (0)475 20 57 67
www.ixor.be

**In our next paper, we'll explore how Agentic AI can earn the trust of users through transparency, predictability, and respectful behaviour. Because no matter how smart an agent is, people won't use it if they don't feel comfortable with it.**