Human + AI Working Together

Microsoft

Chapter 5: Security, Governance and Trust

In Chapter 4, we explored the intelligence layer that powers Microsoft's Agent First vision and examined how organisational context enables agents to deliver meaningful value.

As agents gain access to more information and take on greater responsibility, another question naturally emerges.

How do organisations remain in control?

For many leaders, governance, security and trust are the most important considerations when evaluating the future of AI. The opportunity presented by Human-Agent Teams is significant, but so is the responsibility to ensure these capabilities are deployed securely, ethically and compliantly.

In this chapter, we examine the governance foundations of Agent First organisations and explore how Microsoft technologies such as Entra, Purview, Defender and Responsible AI frameworks help organisations innovate while maintaining trust and accountability.

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

Building Agent First Organisations Responsibly
The Opportunity and the Responsibility

Imagine a future where every employee has access to a team of intelligent agents. Agents that can research information, analyse data, update systems, monitor compliance, coordinate processes and complete tasks on behalf of users. The potential benefits are enormous.

- Greater productivity.
- Improved service delivery.
- Reduced administrative burden.
- Increased organisational capacity.

Yet alongside the opportunity comes an important question: **How do organisations maintain control?**

For many leaders, this is the defining challenge of the Agent First era. The technology itself is exciting. The governance surrounding it is essential.

This is particularly true across government, healthcare, policing and education, where security, privacy and accountability are fundamental requirements rather than optional considerations.

The reality is that organisations will not succeed because they deploy the most agents. They will succeed because they deploy them responsibly. Trust will become one of the most important success factors in the transition towards Human-Agent Teams.

Fortunately, this is an area where Microsoft has significant strengths. The Agent First vision is not being built from scratch. It is being built upon decades of enterprise security, identity management, compliance and governance capabilities that already exist across Microsoft 365, Azure, Dynamics 365, Power Platform and Microsoft Fabric.

The Trust Challenge

Every major technological shift creates new questions.

Cloud computing introduced concerns about data residency and security. Mobile working raised questions about access control and device management. Artificial Intelligence introduces a new set of considerations.

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

For example:

- Can agents access sensitive information?
- What happens if an agent makes a mistake?
- Who remains accountable for decisions?
- How do organisations prevent misuse?
- Can employees trust AI-generated outputs?
- How do we ensure compliance obligations are met?

These are valid concerns. The good news is that most of these challenges are not new. They are extensions of existing governance disciplines.

- Identity.
- Security.
- Data management.
- Compliance.
- Risk management.

The difference is that these disciplines must now be applied to digital workers as well as human workers.

Human Control Remains Essential

One of the biggest misconceptions surrounding Agent First is the idea that organisations are handing complete control to AI. Microsoft's vision does not support that model. Throughout Microsoft's approach, human accountability remains central.

Agents may perform tasks. Agents may coordinate activities. Agents may generate recommendations. But responsibility remains with people.

This principle becomes particularly important when considering higher-risk activities such as:

- Regulatory compliance
- Financial approvals
- Clinical decision support
- Public sector case management
- Operational risk management

In these scenarios, organisations may choose to implement human approval processes before actions are completed. This creates a balanced model. AI provides speed and scale. Humans provide judgement and accountability. The objective is augmentation, not abdication.

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

Identity: The Foundation of Agent Security

One of the most important questions organisations will need to answer is:
"Who is this agent acting as?"

Identity sits at the centre of Microsoft's security model. Just as employees require access rights, permissions and authentication, agents must also operate within clearly defined boundaries. This is where Microsoft Entra becomes critical.

Through Entra, organisations can control:

- Identity
- Authentication
- Authorisation
- Access policies
- Conditional access
- Privileged access controls

This ensures agents only have access to information and systems they are explicitly authorised to use. In many cases, organisations may choose to grant agents less access than human employees. The principle remains the same. Least privilege. Only the access required to perform the intended task. Nothing more.

Permissions Still Matter

A common concern is whether AI can suddenly access information that employees cannot. The answer is **no.**

Microsoft's AI services are generally designed to respect existing security permissions. If an employee does not have access to a document, the AI cannot magically retrieve it on their behalf. The same principle applies to agents.

Agents operate within the permissions assigned to them. This is one of the reasons data governance becomes so important.

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

Poorly managed permissions create risks regardless of whether AI is involved. Agent First simply increases the importance of getting these foundations right.

Organisations preparing for Agent First should ask:

- Do we understand who has access to what?
- Are permissions regularly reviewed?
- Is sensitive information appropriately classified?
- Are data governance processes effective?

The stronger these foundations are today, the smoother the transition to Agent First will be.

Data Security in the Agent First Era

As agents become more capable, data becomes increasingly valuable. After all, intelligence depends upon information. This creates a responsibility to ensure data remains secure.

Microsoft addresses this challenge through multiple layers of protection including:

- **Encryption**
  - Data remains encrypted both at rest and in transit.
- **Identity Controls**
  - Access is governed through established identity management processes.
- **Security Monitoring**
  - Threats and unusual behaviour can be identified through security platforms such as Microsoft Defender.
- **Data Loss Prevention**
  - Policies help prevent inappropriate sharing of sensitive information.
- **Information Protection**
  - Organisations can classify and protect information according to business requirements.

These capabilities already exist today. Agent First does not replace them. It makes them even more important.

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

The Role of Microsoft Purview

As organisations begin deploying agents at scale, governance becomes increasingly important. This is where Microsoft Purview plays a critical role.

Purview helps organisations understand, govern and protect their information. Capabilities include:

- Data classification
- Data lifecycle management
- Compliance monitoring
- Information protection
- Data discovery
- Risk management

For Agent First organisations, Purview provides an essential governance layer. It helps answer questions such as:

- Where is sensitive information stored?
- Who has access to it?
- How is it being used?
- Are compliance requirements being met?

As digital labour expands, visibility becomes increasingly important. Purview helps provide that visibility.

Responsible AI by Design

Microsoft has invested heavily in responsible AI principles. These principles are particularly important as organisations move towards more autonomous systems.

Key areas include:

- **Fairness**
  - AI should operate without creating unintended bias.
- **Reliability and Safety**
  - Systems should perform consistently and predictably.
- **Privacy and Security**
  - Information must remain protected.

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

- **Inclusiveness**
  - Technology should support diverse users and needs.
- **Transparency**
  - Users should understand how AI is being used.
- **Accountability**
  - People remain responsible for outcomes.

These principles are not theoretical. They increasingly influence how AI solutions are designed, deployed and governed. For organisations building Human-Agent Teams, responsible AI should become a core component of governance frameworks.

What Happens If An Agent Gets It Wrong?

Perhaps the most common question leaders ask is:
"What happens when an agent makes a mistake?"

The answer depends on how the organisation has designed the process. Just as human employees can make mistakes, agents can also generate incorrect outputs or take inappropriate actions if controls are not in place. This is why governance matters.

Organisations should consider:

- Human approval checkpoints
- Escalation processes
- Monitoring and auditing
- Testing and validation
- Continuous review

Many organisations will begin with lower-risk scenarios before introducing greater levels of autonomy. For example:

Low Risk	Medium Risk	High Risk
Meeting Summaries	Workflow coordination	Financial actions
Information retrieval	Reporting	Compliance decisions
Research support	Administrative processing	Clinical support
		Public Service case management

Different levels of oversight may be appropriate depending on the scenario. The goal is not to eliminate risk. The goal is to manage it effectively.

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

Building an Agent Governance Framework

As Agent First adoption grows, organisations will increasingly require formal governance structures. A practical governance framework should consider:

- **Strategy**
  - What outcomes are we trying to achieve?
- **Ownership**
  - Who is responsible for agent performance?
- **Security**
  - How is access controlled?
- **Compliance**
  - What regulatory requirements apply?
- **Monitoring**
  - How do we measure effectiveness?
- **Risk Management**
  - What controls are required?
- **Continuous Improvement**
  - How will agents evolve over time?

Organisations that establish these foundations early will be better positioned to scale successfully.

Why Public Sector Organisations Must Lead Responsibly

Across local government, healthcare, policing and education, public trust is essential. Citizens expect services to be secure, transparent and accountable. This means public sector organisations have an opportunity to become leaders in responsible AI adoption.

The challenge is not simply deploying technology.
The challenge is deploying technology in a way that enhances trust.

Done well, Agent First can help organisations improve services, increase efficiency and support employees while maintaining the governance standards citizens expect.

That should be the goal.

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

Key Takeaways

The success of Agent First depends as much on governance as it does on technology. Human oversight remains essential even as agents become more capable.

Identity, permissions and access controls continue to provide the foundation for secure AI adoption.

Microsoft Entra, Purview, Defender and the wider Microsoft security ecosystem play critical roles in supporting Agent First organisations.

Responsible AI principles help ensure that AI remains secure, transparent and accountable.

The organisations that gain the greatest value from digital labour will be those that combine innovation with strong governance, effective security and a culture of trust. Agent First is not about losing control.

It is about creating new forms of capacity while maintaining the controls that matter most.

How CPS Can Help

Successful AI adoption depends on trust. CPS works with organisations to establish governance frameworks, security controls and Responsible AI practices that enable innovation while maintaining compliance, accountability and organisational confidence.

CPS | Microsoft Partner of the Year – Copilot & Agents

AI Strategy | Copilot | Copilot Studio & Agents | Agent Factory | Microsoft Fabric | Dynamics 365 | Power Platform

Get in touch via hello@cps.co.uk

hello@cps.co.uk
+44 (0) 1628-321321
www.cps.co.uk

Copilot Jump Start Partner

Microsoft FastTrack Partner

To find out how we can support your new ways of working, contact us today.

hello@cps.co.uk

+44 (0) 1628-321321

www.cps.co.uk