From Pilot to Enterprise

Why AI succeeds in one department and fails across the enterprise, and how to change that



U-CIE · October 2026

One pilot, six months later

MARCH

~~20 min~~

5 min

to handle a customer complaint with the pilot's AI assistant

OCTOBER

'Is the project working, yes or no?'

Nobody on the leadership team can answer.

March to October: the story of a pilot

March

An AI assistant in customer service. Twenty minutes become five. The head of department wants more.

June

The leadership team approves the roll-out to three more business units. Licences, training, enthusiasm.

September

‘Active customer’ means something different in each subsidiary. Access rights too broad. An out-of-date clause quoted. A connection silent for ten days.

October

‘Is it working, yes or no?’ The first users are delighted. The new business units, much less so.

The AI model didn't change between March and October. The environment did.

The core idea

AI almost never stumbles on its intelligence. It stumbles on the seams of the organisation.

Data that means different things in different departments. Poorly defined access rights. Decisions nobody owns.



A pilot is a dinner for six, not 300 covers



In the pilot

- A use case chosen because it is simple
- Documents sorted by hand
- An expert within earshot
- Errors with little consequence





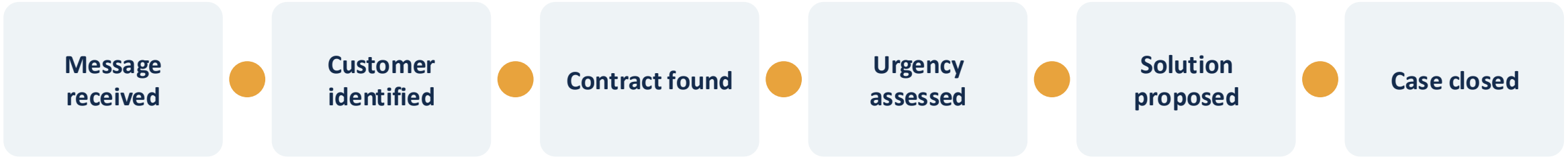
Across the enterprise

- Varied cases, some highly unusual
- Live data to maintain over time
- Users who don't know whom to call
- Errors that affect a customer, a payment, an employee

A pilot proves usefulness. It does not prove that the organisation can run the use without its protections.

The race is lost at the baton changeover

Example: the real flow of a customer complaint, from message received to case closed



Each gold dot is a link. For the next step to work, something has to pass through it correctly:

- a piece of data
- a right
- a decision
- a responsibility

Plugging in an AI system resolves none of these links.

Six links to check in every flow

Data

What do we mean by an ‘active customer’, and which system is the reference?

Applications

If this connection fails, who notices, and how quickly?

Identity and rights

Can the AI see or do something the user is not allowed to?

Decision

Who decides, and by what rule, when a case is out of the ordinary?

Accountability

Who has the right to press the stop button?

Measurement

What was our starting figure, and what do we review each month?

September, reread through the six links

‘Active customer’ means something different in the subsidiary and at head office	**Data**
An adviser can see information he never had access to before	**Identity and rights**
A connection broken for ten days without anyone noticing	**Applications and accountability**
In October, nobody can say whether the project works	**Measurement**

Each September incident is a link nobody had dealt with. None of them is an AI failure.

AI speeds everything up, including errors

Five cascades: the output of one step becomes the action of the next, and the error grows along the way

1

Data

A wrong customer reference, confidently reworded, attached to the wrong account.

2

Decision

A wrong classification becomes a wrong allocation, then an unhappy customer.

3

Rights

A booby-trapped email gives orders to an assistant with access that is too broad.

4

Confidence

A wrong but credible answer moves from note to report, then into a decision.

5

Accountability

Everyone sees one piece of the system. Nobody can stop it.

The right question

~~Not: ‘Can we trust AI?’~~

But: in this flow, which AI outputs do we give what degree of trust, to trigger which action, under what control and with what recourse?

Four levels of autonomy, set task by task

More autonomy, more evidence required

1

Suggest

It informs a decision

2

Prepare

It prepares an act; a person approves it

3

Execute

It carries out a pre-written scenario on its own

4

Orchestrate

It chooses for itself the sequence of several authorised actions

The same assistant may suggest for one task and execute for another. You set each task, not the tool.

Before granting autonomy, three verbs



Understand

Know what the AI can do, and with what access



Monitor

Know where to observe its actions, and who receives the alerts



Contain

Be able to restrict it, pause it, stop it

'If the organisation cannot understand, monitor or contain an agent's actions, it is not ready to be deployed.'

Test proposed by the NCSC, the UK's National Cyber Security Centre (paraphrased)

Seven workstreams, none replaces the others

1

Design the flow before choosing the tool

2

Treat data as a shared product

3

Separate understanding from acting

4

Name who decides what

5

Evaluate the system, not just the model

6

Train for the flow, not the prompt

7

Manage a portfolio, not prototypes

The weakest workstream sets the level of autonomy you can reach.

EU law judges the use, not the tool



Beware the optical effect of the postponement. It concerns only high-risk systems. The GDPR, AI literacy, transparency and your security obligations continue to apply.

Legal position on 25 October 2026, after Regulation (EU) 2026/1744. This material is not legal advice.

Five sectors, the same mechanism

Hélianthe Exploitation

Renewable energy

Data

A good summary does not say which definition is authoritative.

Signal Urbain

Field marketing

Data

AI formats information. It does not turn an estimate into proof.

Club Partenaires Europe

Sports hospitality

Rights

Preparing an offer is not the right to commit to it.

NordVale Finance

Lending

Decision

Summarising a file is not judging a file.

Mécaforge Europe

Manufacturing

Decision

A recommendation must not drift into an order.

In all five cases, the AI works. It is a link that breaks.

What you are really paying for

What you see: platform, licences, connectors

What the demonstration does not show



What share of the budget funds the tool, and what share funds the links that will make it reliable?

A budget that funds only the tool is really funding one more pilot.

Five signs that a project is not ready

- Nobody can say who has the right to stop the system
- The assistant uses an account with broader rights than its users
- Success is measured only by the satisfaction of the first users
- Exceptions are ‘handled by the project team’ without being written down
- Nobody can reconstruct what the AI read and triggered last week

A warning sign?

It is not a refusal. It is a workstream to fund before scaling up.

The white paper also offers ten questions to tick off in the meeting.

Two hours to map three flows

1 15 min

Framing

Choose three flows with different consequences

2 30 min

The real flow

The person who does the work describes a recent case

3 30 min

Spotting the links

Fill in the map, writing ‘unknown’ freely

4 20 min

Cascade simulation

If the AI gets it wrong here, how far does the error go?

5 15 min

Levels of autonomy

Set the reasonable level, task by task

6 10 min

Decisions

Three workstreams, three owners, one date



Share of the 120 minutes. Around the table: chief executive, finance director, IT director, business owners, and someone who does the work day to day.

Which flows do we want AI to make more capable, and which links must we strengthen so that this capability remains trustworthy at scale?



U-CIE

Think through, build and steer your AI, from A to Z



Think

Map with your teams the flows that matter and the links to strengthen



Build

Design your AI uses and set out, for each, what it may do on its own



Steer

Measure value, monitor risks, adjust autonomy

AI should transform your organisation without ever slipping out of your control.

u-cie.com