

Same Prompt, Three Model Modes

A teaching artefact for comparing Instant, Thinking, and Pro responses to one AI project-prioritisation task.

Core lesson

Model mode affects speed, depth, risk sensitivity, and the recommendation a person receives.

The Experiment

Prompt used in all three runs

A company wants to introduce AI and automation into its operations.
They have identified five possible projects:

1. Automated meeting summaries - Summarise internal meetings and extract actions.
2. Customer email triage - Categorise incoming customer emails by urgency and topic, then route them to the right team.
3. AI sales proposal generator - Generate first drafts of sales proposals using CRM notes and previous proposal templates.
4. Invoice processing automation - Extract data from supplier invoices and enter it into the finance system.
5. Employee HR chatbot - Answer common HR policy questions using internal documents.

The business has limited time and budget.

Rank these ideas from highest to lowest priority using four criteria: Business value, Ease of implementation, Risk, Measurability.

Recommend which project should be done first and explain why.

How to read the evidence

- Hold the prompt constant. The useful comparison is what each model mode chooses to emphasise.
- Look for the scoring method, not just the final ranking.
- Notice which risks are treated as manageable and which risks change the recommendation.
- Ask what a human decision-maker still needs to specify before accepting the ranking.

Human judgement checkpoint

The prompt does not state which criterion matters most. If the organisation cares most about fast customer impact, customer email triage may win. If it cares most about controllable ROI, invoice processing may win.

WHAT CHANGED ACROSS MODES

At-a-Glance Comparison

Mode	First recommendation	Ranking order	Scoring style	Visible thinking time	Teaching takeaway
Instant	Customer email triage	Email triage; Invoice automation; Meeting summaries; Sales proposal; HR chatbot	Numeric 1-5 scores with a total out of 20.	Immediate / not shown	Fast, structured, and useful for a quick first pass. It gives a plausible answer cleanly, but does less visible trade-off testing.
Thinking	Customer email triage	Email triage; Invoice automation; Meeting summaries; HR chatbot; Sales proposal	Qualitative ratings: High, Medium, Low, plus an overall view.	A couple of seconds	Adds caution and control language. It keeps the same top recommendation, but treats sales proposal generation as riskier than HR chatbot.
Pro	Invoice processing automation	Invoice automation; Email triage; Meeting summaries; Sales proposal; HR chatbot	Numeric scores, but framed around risk control and mature implementation.	50 seconds	Changes the recommendation. It gives more weight to measurable ROI, mature technology, and controllable back-office risk.

What this shows

The models were not solving a maths problem with one fixed answer. They were applying judgement to underspecified criteria. The important learning is to inspect the weighting behind the answer, especially when risk and measurability can be interpreted in more than one way.

Instant Evidence Page

A simple and effective way to approach this is to score each idea against the four criteria on a 1–5 scale (5 = best). For **Risk**, a higher score means **lower implementation risk**.

Project	Business Value	Ease of Implementation	Low Risk	Measurability	Total
2. Customer email triage	5	4	4	5	18
4. Invoice processing automation	5	3	4	5	17
1. Automated meeting summaries	3	5	5	3	16
3. AI sales proposal generator	4	3	3	3	13
5. Employee HR chatbot	3	2	2	2	9

Ranking

1. Customer email triage ★ (Recommended First Project)

Why it ranks first

- **High business value:** Every customer enquiry is handled faster and more consistently, improving

+ Ask anything Instant ▾  

Screenshot evidence from the model response.

What to notice

- Uses a 1-5 scoring scale and adds totals, making the answer easy to scan.
- Recommends customer email triage because it combines high value, low operational risk, and clear metrics.
- Keeps the rationale broad and confident. It gives less attention to what could go wrong if urgent messages are misrouted.

Response notes

"Customer email triage" is recommended as the first project because it offers high business value, moderate implementation effort, low operational risk, and clear measurable outcomes.

Teaching notes

Instant is useful when people need a quick shortlist or a starting hypothesis. It should not be treated as the final decision when the criteria are underspecified.

MORE EXPLICIT RISK LANGUAGE WHILE KEEPING THE SAME TOP CHOICE

Thinking Evidence Page

Thought for a couple of seconds >

Here's a practical ranking using the four criteria.

Rank	Project	Business value	Ease	Risk	Measurability	Overall view
1	Customer email triage	High	Medium	Medium	High	Strong operational value, easy to measure, directly improves customer response time
2	Invoice processing automation	High	Medium	Medium-High	High	Clear time and cost savings, but needs stronger accuracy controls
3	Automated meeting summaries	Medium	High	Low-Medium	Medium	Easy pilot, low risk, useful productivity gain, but value may be diffuse
4	Employee HR chatbot	Medium	Medium	Medium-High	Medium	Useful for reducing repeat HR queries, but policy accuracy and trust matter
5	AI sales proposal	High potential	Medium-Low	High	Medium	Could create revenue value, but quality, brand risk and CRM/template readiness

+ Ask anything

Thinking ▾

Screenshot evidence from the model response.

What to notice

- Uses qualitative ratings rather than numeric totals, which makes trade-offs easier to discuss.
- Still recommends customer email triage, but says it should route and recommend rather than send final responses automatically.
- Ranks HR chatbot above sales proposal generation because sales has higher brand, quality, and data-readiness risk.

Response notes

Customer email triage has the best balance across the criteria, but the model adds that the system should route and recommend while humans retain review or override control.

Teaching notes

Thinking mode is useful when a comparison needs more caution, especially where the answer depends on controls, governance, and rollout design.

Pro Evidence Page

Thought for 50s >

Recommended priority ranking

I would score each project on a 1–5 scale, where 5 is best. For risk, a higher score means **lower risk** / easier to control.

Rank	Project	Business value	Ease of implementation	Risk control	Measurability	Overall view
1	Invoice processing automation	5	4	4	5	Best first choice: clear ROI, mature technology, easy to measure
2	Customer email triage	5	3	3	5	High value, but needs careful controls to avoid misrouting urgent issues
3	Automated meeting summaries	3	5	4	3	Very easy quick win, but business value may be more diffuse
4	AI sales proposal generator	4	3	3	3	Potentially valuable, but quality, brand, and approval risks are higher

+ Ask anything

Pro   Useful, but policy accuracy

Screenshot evidence from the model response.

What to notice

- Changes the top recommendation from customer email triage to invoice processing automation.
- Weights mature technology, measurable cost savings, and back-office controllability more heavily.
- Treats customer-facing mistakes as more operationally sensitive than finance extraction errors with human review.

Response notes

Invoice processing automation is recommended first because it has clear ROI, mature OCR and workflow technologies, measurable costs, and manageable risk through exception handling and human review.

Teaching notes

Pro is useful when a decision deserves deeper scrutiny. The trade-off is speed: this run visibly took 50 seconds, so it should be reserved for tasks where the extra judgement is worth the wait.

Reflection

Discussion prompts

- Which output would you use as a first draft, and which would you challenge?
- Which criterion did each mode appear to weight most heavily?
- If your organisation had limited budget, would you optimise for fast visible impact or controllable ROI?
- What missing information would you ask the business for before accepting any ranking?

When each mode is good enough

Instant: quick shortlist, rough prioritisation, first-pass explanation.

Thinking: trade-off discussion, implementation planning, risk-sensitive comparison.

Pro: deeper review, high-impact choices, situations where the first plausible answer may be misleading.

Human judgement checkpoint

Before acting on a model ranking, define how the four criteria are weighted. A customer-service leader might prioritise customer email triage. A finance or operations leader might prioritise invoice automation. The model can structure the decision, but the business must own the weighting.

Takeaway

Reasoning models do not remove judgement. They make more of the judgement visible, and sometimes they surface a better challenge to the obvious answer.

Suggested classroom activity

Ask participants to defend one of the three recommendations in pairs. One person argues for customer email triage, one argues for invoice processing automation. They must use the four criteria and name the extra business information they would need to decide.