



# Summer School 2025

## Pitfalls in the evaluation process and how to avoid them

---

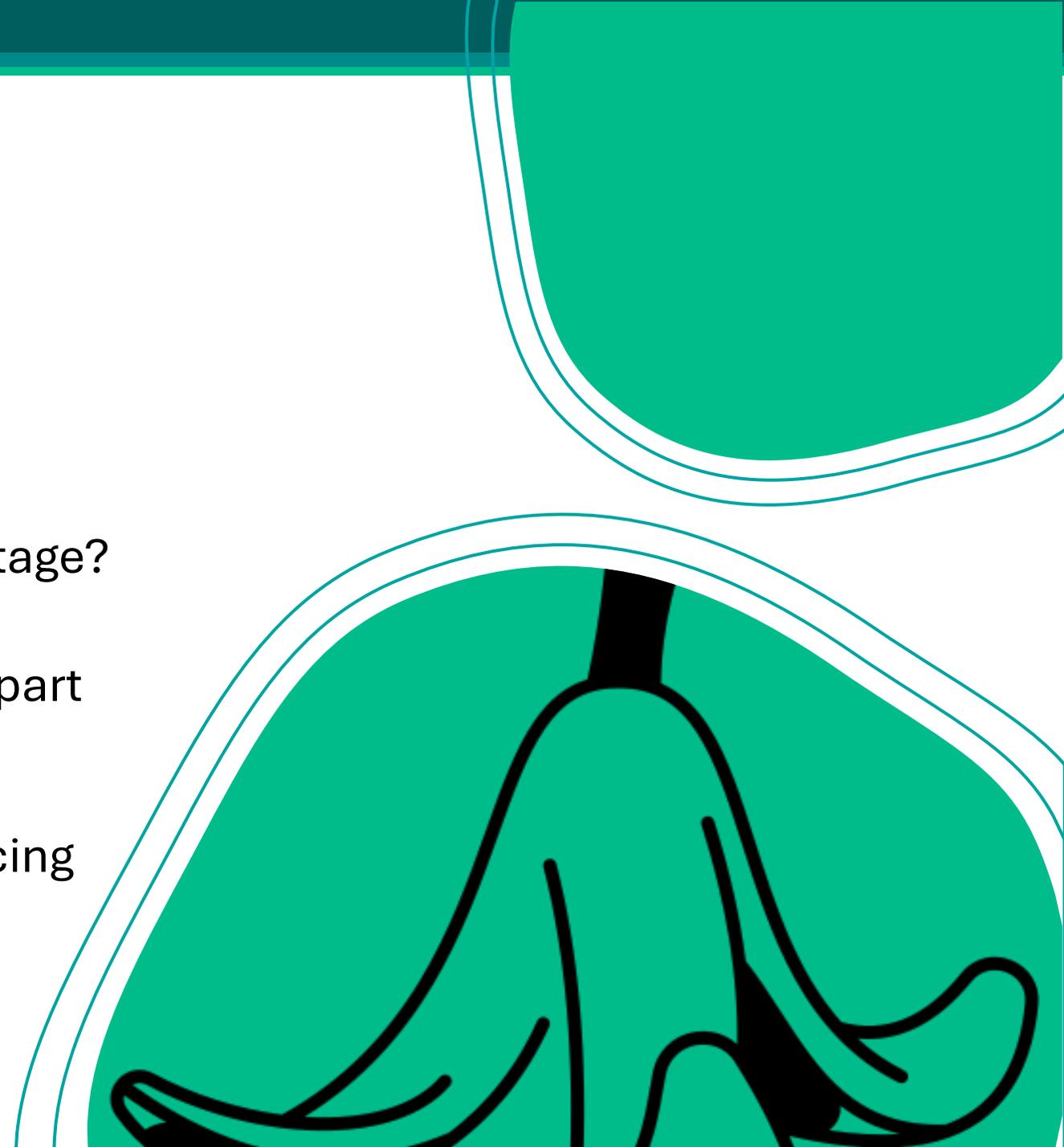
European Centre of Solidarity  
Gdańsk, Poland  
(2-4 July 2025)



# Day 3

# AI and pitfalls

- What are your reflections at this stage?
- How is AI part of the problem? Or part of the solution?
- What are the problems you are facing when trying to implement AI solutions?



# Guiding principles for working with an AI

1. Only use an LLM for production **if you can control** the quality of the process and results
2. For simple tasks, **choose more robust & auditable solutions** (e.g. Excel)
3. Cut complex tasks into **smaller ones** and identify which can be performed by AI, and which by humans
4. Be very **careful about using AI** with private or sensitive data or with vulnerable groups
5. Be aware of **cultural gaps** in input (what a U.S. AI can't grasp) and output (how awkward?)
6. Develop **AI skills**, use AI for tasks for which you or colleagues are skilled
7. Use AI process for **reflexivity** about questions asked and answers received
8. All uses of AI should be **disclosed**

# Agree on uses and risks in ToR / inception phase

Translating sources, summarising, structuring, increase readability, writing excel formulas or R programme, contacting actors, narration, visualisation, multimedia production  
...

Declaration Only

**SAFE**

Contributing to development of ToC; high-level coding of documentation or portfolio; drawing 'causal landscape'; AI interviewing; coding of simple interviews; cross-analysing / clustering datasets; testing traceability of findings; manage evidence (RAG); summarising findings, ...

Specific protocol and oversight, plan B

**CAUTION**

Any use with sensitive / confidential data, Review literature\*, formulate EQs, produce research design, create protocols or questionnaires, sample cases, interpret data or make judgments,  
...

Make collaborative

**AVOID**

# AI reveals problems with evaluation practice

Evaluation Unit officer  
working alone

Evaluation limited to  
descriptive activity

Lack of Quality  
Assurance process

Insufficient skills /  
knowledge  
management

# Future of AI in evaluation?

Regulation and Monetisation of AI Use

In-house AIs (LLM + RAG) for knowledge management

Automated M&E systems

Standardisation of 'routine' evaluation tasks & Agentic AI

Pre-packaged 'AI-driven evaluation' offers

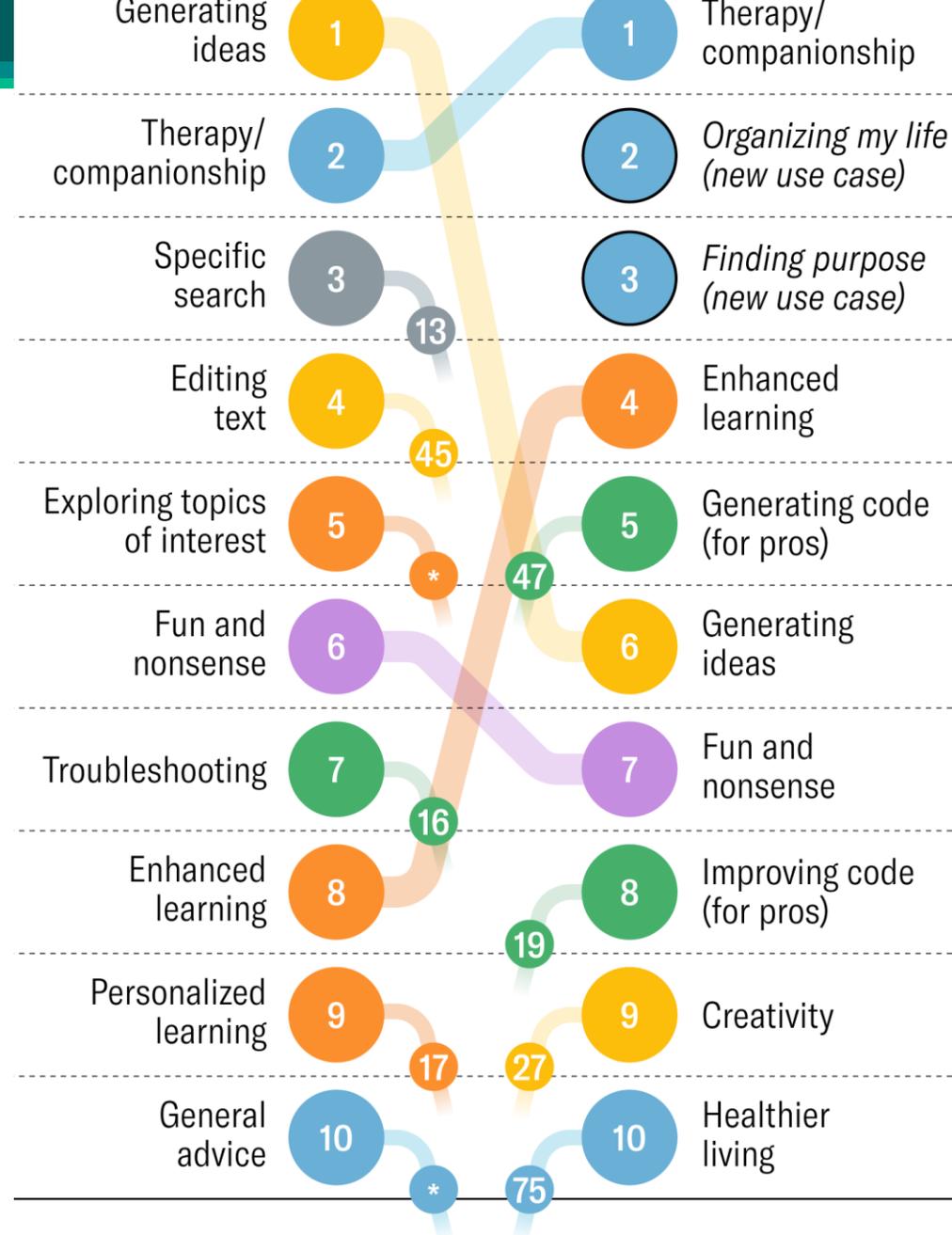
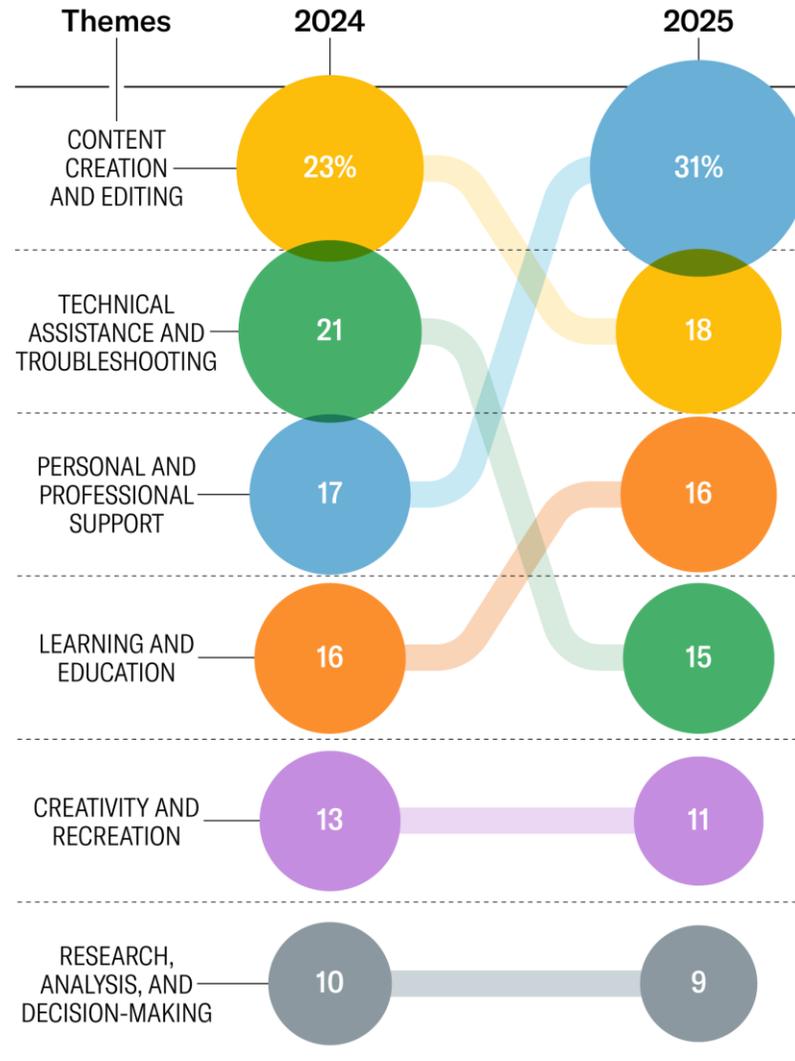
...

# Major Gen AI Use Case Themes That Emerged

Use cases have shifted from technical to emotive over the past year.

## Uses are changing fast

Zao-Sanders, M. (2025). How people are really using Gen AI in 2025. Harvard Business Review. <https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025>



# A systemic view about pitfalls

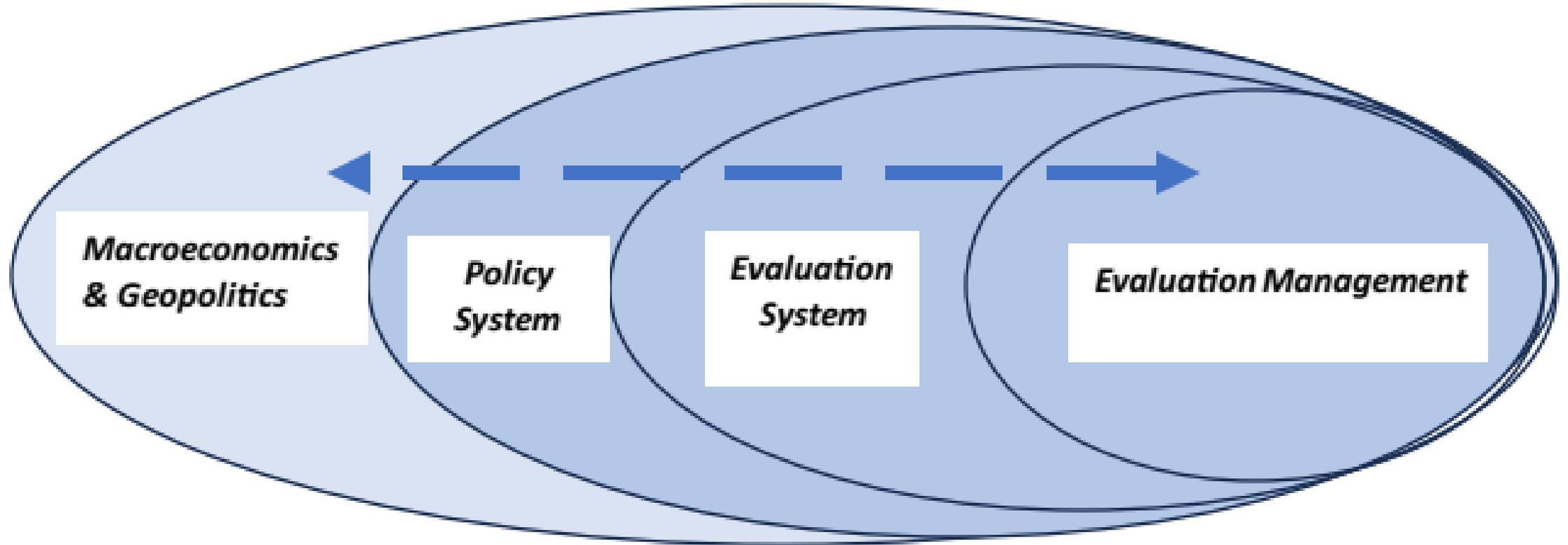


- Evaluation is never self contained
- Rather, it is embedded in wider systems:
  - Administrative systems with rules and resources
  - Evaluation systems with knowhow, networks & institutions
  - Policy systems with priorities and interests – that can change while we are evaluating!
  - Geopolitical systems that bring us wars, climate change and pandemics

# A 'systemic' view makes it easier to remedy evaluation weaknesses

- Engaging with stakeholders across the evaluation cycle increases the likelihood of evaluation relevance, learning and use
- Better planning and prioritisation can ensure that scarce evaluation resources are well-targeted
- Smart specifications compensate for limitations in evaluation supply (e.g. *Evaluability Assessments that prepare ToRs; Approach Papers that pre-prepare ToCs; staged evaluations to ensure flexibility; commissioning 'Knowledge Reviews' from local universities*)
- Ongoing evaluation network-building increases likelihood of better bids when needed
- Joint evaluations across different programmes and even with other MAs can increase evaluation efficiency and make more sophisticated methods more affordable.

Evaluation Management is embedded in wider systems...  
which shape evaluations and to an extent are also shaped by  
evaluation



# Systemic problems emanating from evaluation system

- Underdeveloped evaluation market– limited supply
- Lack of methodological knowledge/experience among potential evaluators e.g. ToCs, RCTs and quasi experimental methods, case studies, synthesis studies...
- Poor project management capability among potential contractors (e.g. timetabling, risk management)
- Non-participation/involvement in evaluations by University sector

*...Need for longer-term investment in evaluation system capacity development*

# Systemic problems emanating from policy systems

- Poorly developed data/indicator systems
- Low legitimacy of public actors – fear of criticism, reputational risk, inflexible procurement arrangements
- Limited experience of ministries and institutions to cooperate with each other
- Political instability undermines initiative scope of civil service and public action

*... Need for investment in data/knowledge systems; adaptive planning; and adaptive policy-making*

# Systemic problems emanating from global uncertainties

- Changes in policy and programme contexts
- Rationale for evaluations may disappear once evaluation underway
- Need to redirect resources to new priorities mid term
- Maintaining strategic priorities while changing the ‘vehicle’
- Need for consensus infrastructure with stakeholders and civil society
- ‘Business Continuity’ modelling – lessons from Covid & virus hacks

*... Need for evaluation managers to plan for uncertainty/crises/risks and contingencies*

Many problems/pitfalls cross system boundaries. For example: ¶

Low-capacity *evaluation systems* >>>>>> ¶

Limited evaluation knowhow >>>>>>>>>> ¶

Limited contractor choice >>>>>>>>>> ¶

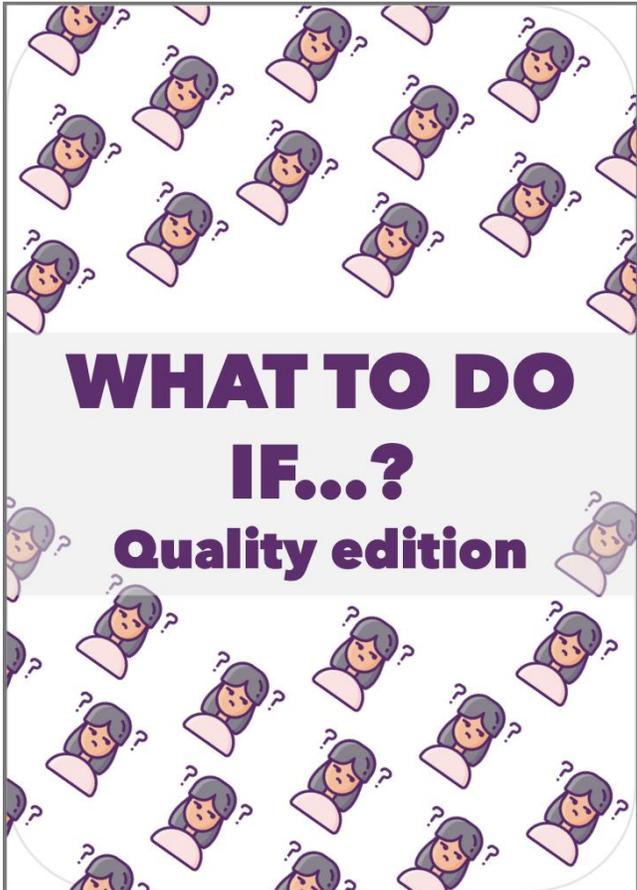
Poor methods application >>>>>>>>>> ¶

Poor evaluation quality >>>>>> ¶





# Wrap-up



# WHAT TO DO IF...?

## Quality edition

### WHAT IF...?

"I have received the draft final report but it doesn't answer the evaluation questions."



### WHAT IF...?

"The findings in the draft final report are not always sufficiently supported by evidence. Some of them look like expert statements on the evaluated issue."

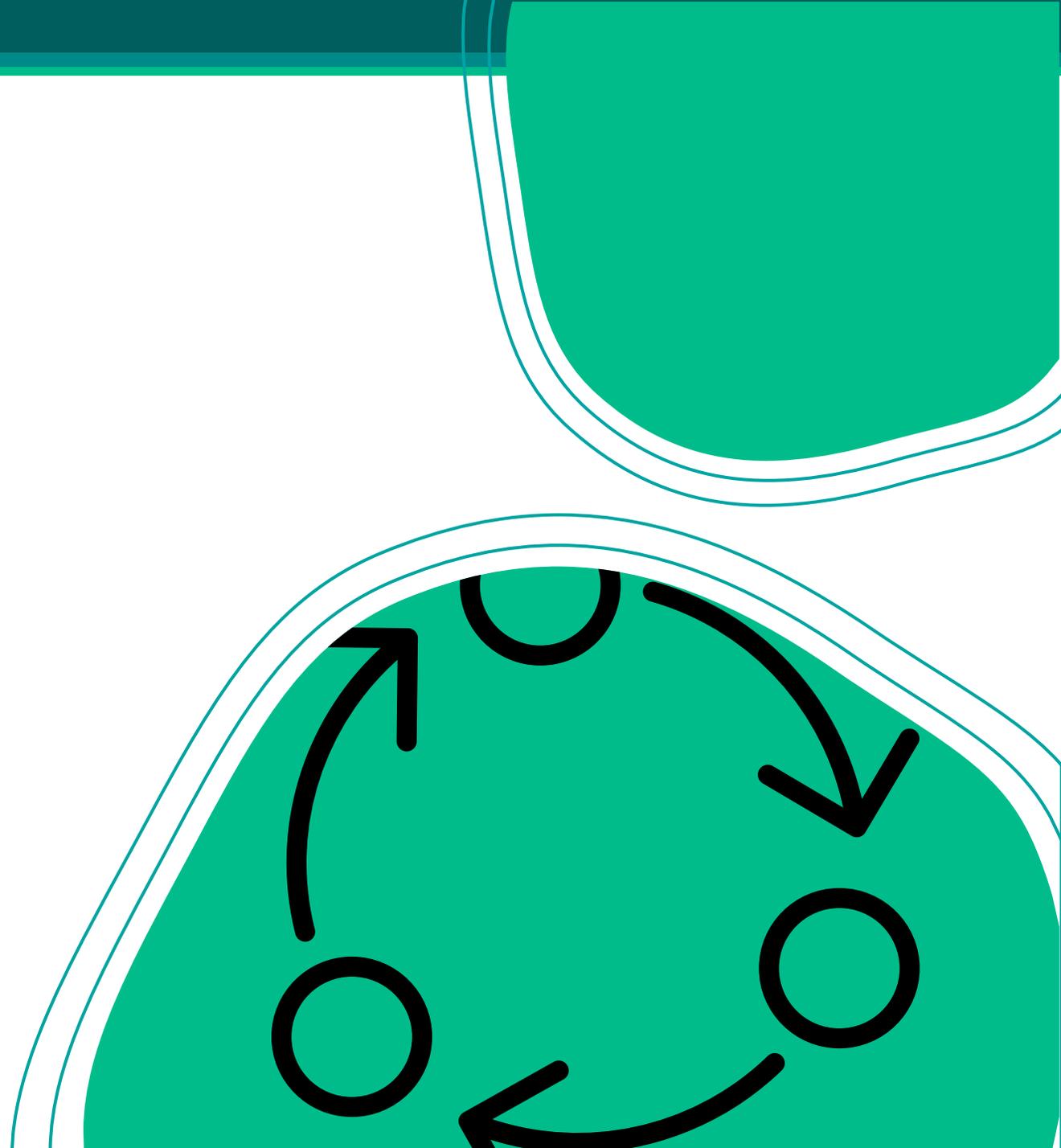


### WHAT IF...?

"Colleagues in the administration question the quality of the final report, but it's really because they don't agree with one specific finding."



# Masterclass





Evaluation Helpdesk on  
Cohesion Policy

[https://www.linkedin.com  
/groups/14420317/](https://www.linkedin.com/groups/14420317/)



**Evaluation Helpdesk (2025)**

