Day 2

2

Evaluation Helpdesk on
Cohesion Policy

https://www.linkedin.com
/groups/14420317/

Problems and Pitfalls - Implementation

- "Data not available when requested by evaluators"
- " Low response rate surveys, data validity risks "
- " Poor analysis, limited methods applied , not integrating statistical and qualitative "
- " Consultants Knowledge gaps e.g. not knowing about methods they put forward in proposals"
- " Poor project management by consultants – effects coverage and timing of what is covered "
- " MAs overloaded by coincidence of activities at the same time – leading to management overload"
- " Unwillingness of contractors to revise at Inception because of contracts"

What participants said…

Many pitfalls along the way…
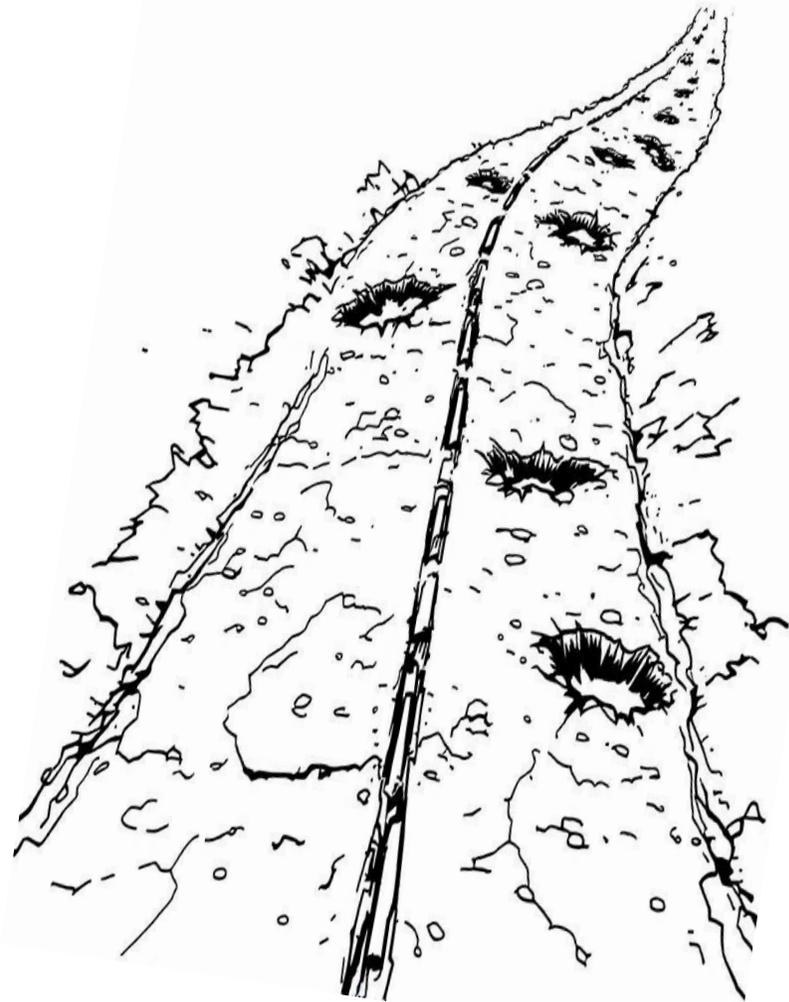
- Consultant doesn't seem to collect information along the EQs/the theory agreed upon in the inception phase
- Fiscal, Social, Monitoring… data never arrived or unusable to answer the questions
- Despite initial promises, competence / profiles that are needed are missing - Principal consultant is not working on the project enough
- Key stakeholders / beneficiaries refuse to answer the questions of the evaluation team
- It is increasingly clear that a course change is necessary to gather sufficient evidence and corroborate findings, but the consultant is unwilling to alter the approach
- Potential users don't seem interested with the emerging results
- …

APPLICA

ISMERI EUROPA

QUADRANT CONSEIL
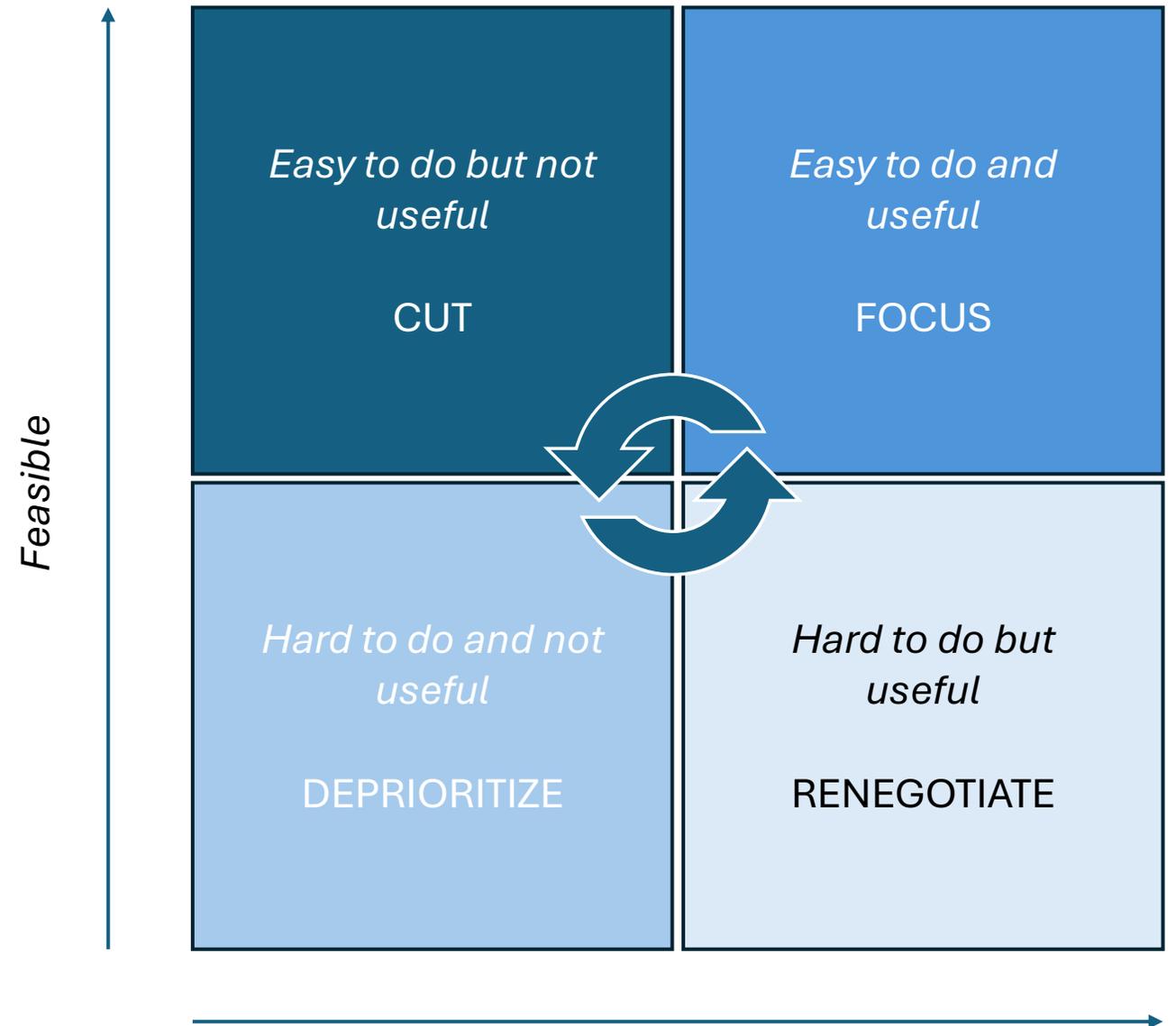
Many possible reasons for this...

- Usually this is collective issue, and sometimes it is nobody's fault

- For example, the competitive process of bidding for evaluations can encourage over-promising of different kinds. Service providers promise to...
  - do more than the budget allows
  - to cover bigger samples than is practical in the timescale
  - to use methods that proposers only half understood but perhaps are eloquently described in proposals - with the help of AI

APPLICA ISMERI EUROPA QUADRANT CONSEIL

If an over-promising evaluation has been accepted and this becomes obvious...

- ... the usual response is to combine contract enforcement with a more pragmatic acceptance that less will in fact be done

- Deciding what aspect of an evaluation can be cut with minimum damage requires judgment in context.

- Which cuts will:
  - upset fewest stakeholders,
  - least weaken core analyses
  - be least likely to stop an evaluation answering priority Evaluation Questions

Use Inception phase to adjust

- The inception report should not duplicate the proposal !

- Use this phase to check feasibility and usefulness e.g. in answering E.Qs

- Renegotiate within boundaries of data, time and budget

- You may need to do that several times in complex evaluation settings

Feasible

Easy to do but not useful	*Easy to do and useful*
CUT	FOCUS
Hard to do and not useful	*Hard to do but useful*
DEPRIORITIZE	RENEGOTIATE

Useful

APPLICA ISMERI EUROPA QUADRANT CONSEIL

How to prevent this sort of situation from happening again?

- Undertaking an evaluability assessment before drawing up a ToR

- Requiring proposers to conduct their own 'risk analysis' including indications of how these risks will be mitigated if they occur – and requiring that such risk analyses are annexed to proposals

- Including in the proposal assessment process, time and budget 'costings' so as to check how realistic are assumptions made

- Building a database across many evaluations, of costs and timings such that over time proposal metrics can be compared against established norms

- Establishing a panel of experts knowledgeable about specialist methods – QCA, Quasi-experiments, cost-benefit analysis, Realist Synthesis, Contribution Analysis, Bayesian updating etc

- Ensuring that contracts allow remedial action if proposals are not followed

APPLICA

ISMERI EUROPA

QUADRANT CONSEIL

Consult with stakeholders all along the evaluation

- Many evaluators focus consultation with stakeholders at the design and delivery stage of an evaluation.

- But stakeholder interests can evolve whilst an evaluation is taking place and evaluations also evolve during implementation

- Maintaining a dialogue with key stakeholders (including partners and beneficiaries) during implementation can be an important 'protective' measure.

APPLICA   ISMERI EUROPA   QUADRANT CONSEIL

Require regular reporting

- Inserting key 'milestones' into ToRs and Contracts so as to identify 'risk factors'

- Contractual reporting more frequently than substantive reporting (e.g. reports on findings) covering progress with fieldwork and data collection in relation to identified milestones

- In addition, methodological and progress reviews can be specified as an essential component of an inception report and draft interim and final reports.

Some potential uses of AI in the design of the approach

- Broaden range of data sources and methods that can be used:
  - Text mining / Coding of reports, forms, etc.
  - Network analysis
  - Clustering, modelling
  - Identification of conditions, risks, enablers (that can be later investigated)

  - ⚠️ Use AI to write code, avoid using LLM directly

- What should be done with caution or avoided?

Agree on AI use in implementation

- Discuss AI uses foreseen or emerging from evaluation process
  - Use opportunity to engage in discussion about protocols and data
  - Engage stakeholders (data party!)
- Different tools:
  - AI priorisation: Cheap, Fast, or Good (no, you can't have the three)
  - Safe / Caution / Avoid framework
  - AI log

- Use AI to discuss difficult questions, e.g. traceability of findings, clarity

WG3

- What you think are the **best** options to provide remedies in the short run or at least to minimise downsides for the SuCIP evaluation of these pitfalls. (You do not need to restrict yourself to the options identified in this scenario.)

- What you think you might do differently in the future to both *anticipate* such pitfalls and be better able to *implement short term remedies* which at present seem to be limited.
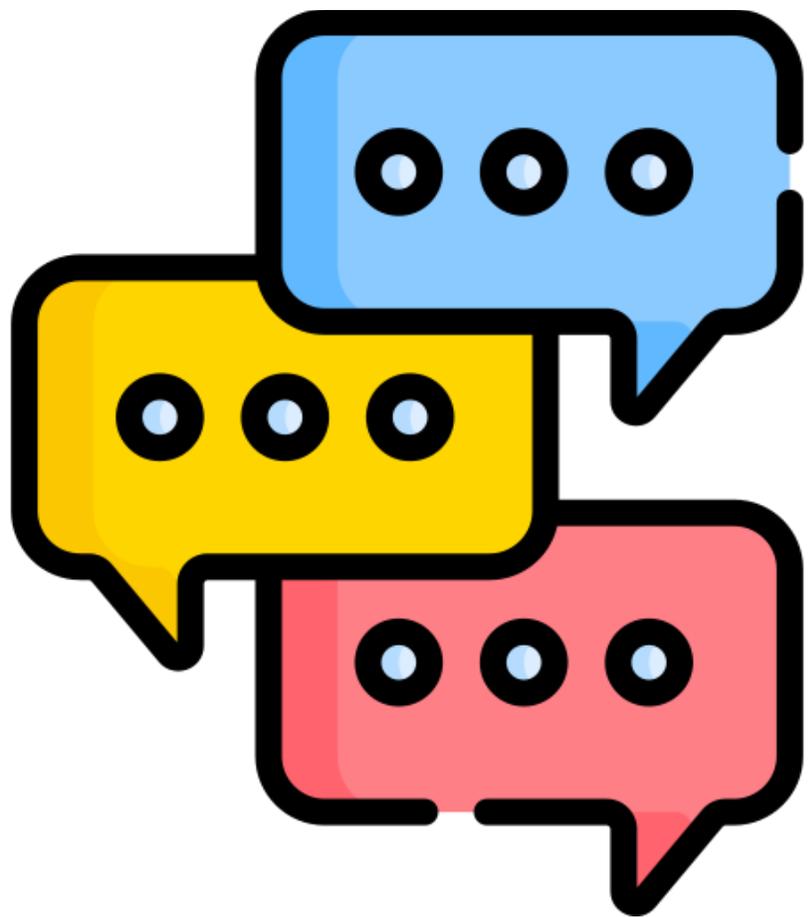
Panel discussion

Pitfalls in the delivery stage
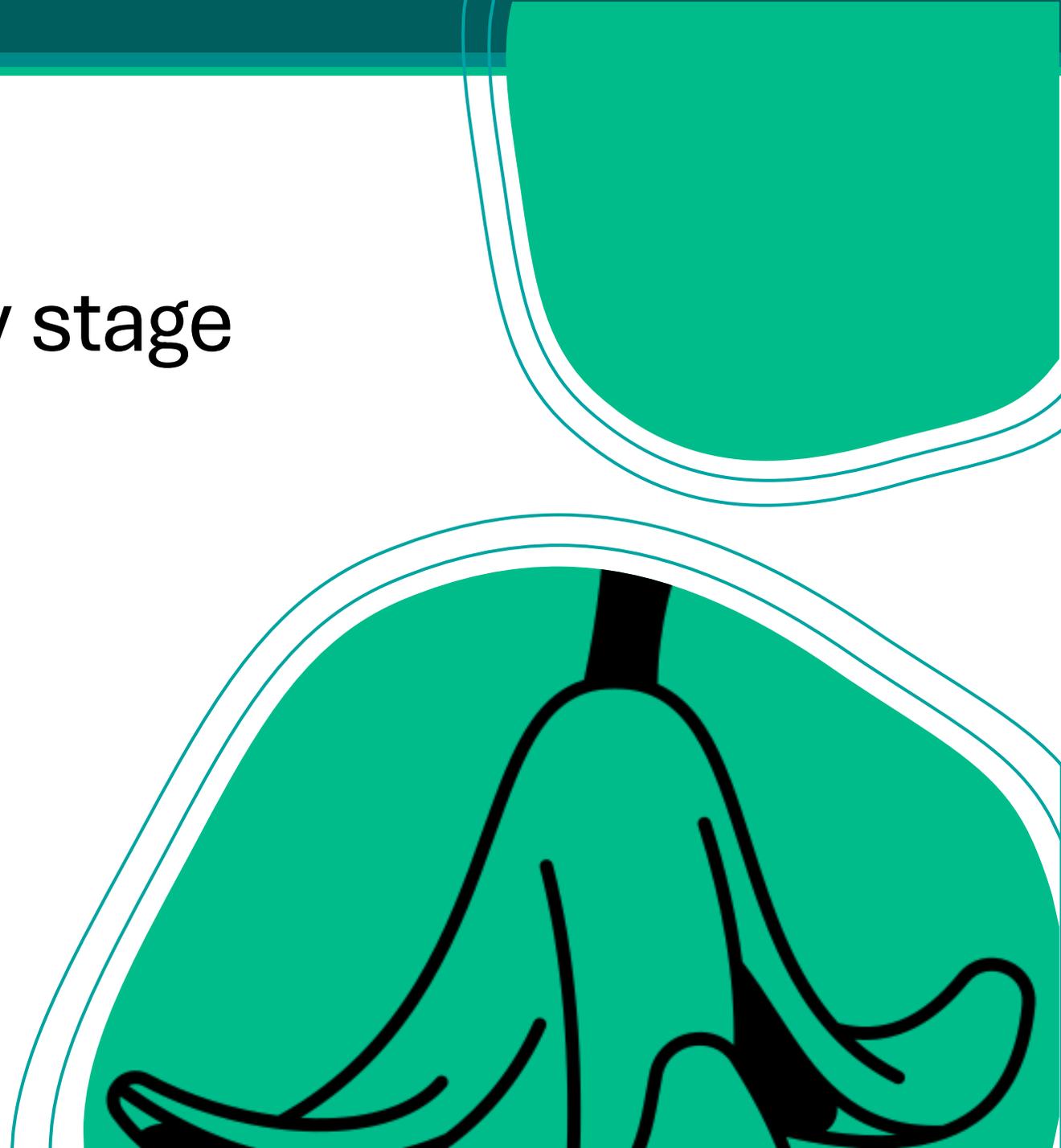
Problems and Pitfalls - Delivery

- "Poor quality reports need redrafting"

- "Evaluation Questions not answered"

- "Insufficient questioning of intervention design itself - rather than how well implemented"

- "Recommendations poorly specified – too general, not context specific, not operational/practical, not linked to evidence"

- "Convincing decision makers and administrative units to use evaluation results"

- "Recommendations to external entities and multiple institutions are most difficult to implement"

- "Policy makers don't think evaluation answers their evaluation questions"

- "Political sensitivity if evaluation is critical of public/regional /local administrations. Can threaten publication of reports"

What participants said…

APPLICA

ISMERI EUROPA

QUADRANT CONSEIL

Keywords when thinking about 'evaluation delivery'

Reporting

Using

Reusing

Learning

Dissemination

What do evaluation delivery 'pitfalls' look like?

Delivery happens on receiving a final report - What about interim reports, other outputs? – seminars, knowledge reviews, networks – also need to be captured

Delivery is when stakeholders 'learn' -receive a report answering their evaluation questions: If EQs not answered is this a 'delivery' failure?

Must an evaluation report be high quality to say it was 'delivered'? Probably a minimum condition although we also learn from failure…

What about longer-term users, those who consult a report in a few years time – perhaps as part of a synthesis review? If not possible equals 'failure'?

Who is responsible for elapsed 'dissemination'? Is non dissemination in the future a delivery failure now?

APPLICA   ISMERI EUROPA   QUADRANT CONSEIL

A tentative definition of evaluation delivery

Deliverables of adequate quality that are produced at different stages of an evaluation process

Which answer pre-specified EQs but also support emergent and unanticipated learning needs

Which are disseminated and targeted to different potential users – policy makers, future policy planners, parliamentarians, citizens

That are stored/remembered in institutional/organisational memories, accessible in both the short and long term

That can be re-analysed, aggregated & recombined; stored and retrieved….. applause from AI aficionados

The centrality of an evaluation 'Delivery Plan'

Given the diversity of….

- Evaluations
- Stakeholders and potential users,
- The different moments that uses and reanalyses of evaluation results may occur,
- Ways in which data, analyses, and derived conclusions may need to be revisited….and stored…

Any single way of specifying delivery risks failure –anticipation and prevention requires configurable 'Delivery Plans'

APPLICA

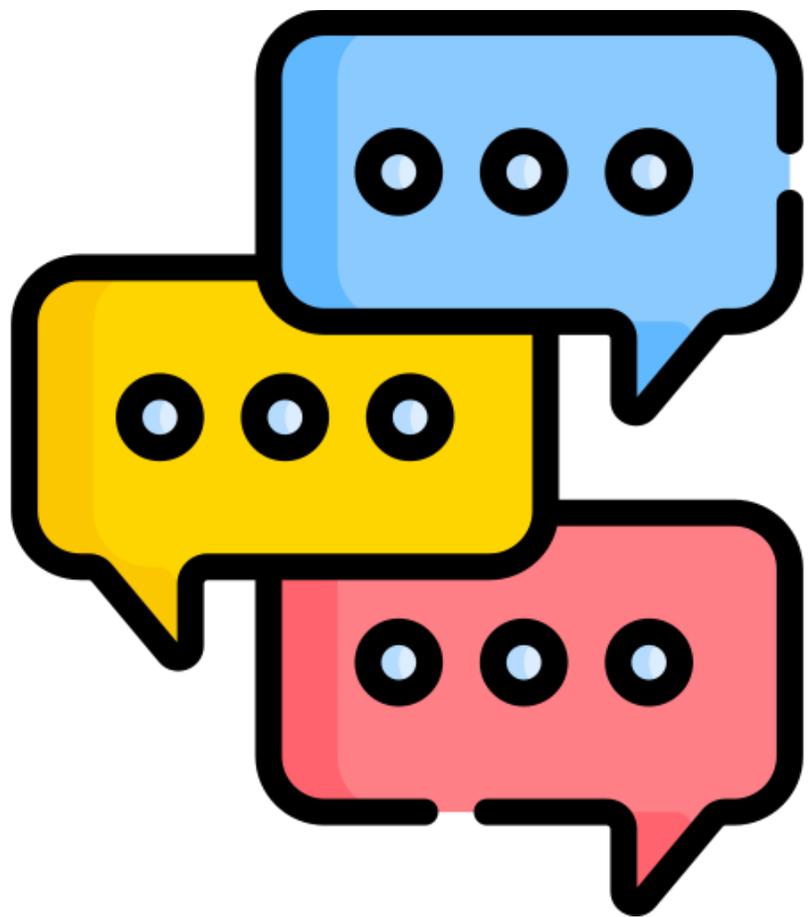ISMERI EUROPA

QUADRANT CONSEIL

Some potential uses of AI in delivery

- Trace finding robustness / Analyse report with different agents

- Extract key points and check whether they match conclusions and recommendations

- Try different summaries by / for different actors

- Disseminate content with podcasts, data viz, video, etc.

- What should be done with caution or avoided?

WG4

- What should the template for an Evaluation Delivery Plan include?

- How would such a plan support future re-analysis, recombination, and accessing evaluation results and deliverables?

- How can the template capture soft outcomes, including the experience and lessons learned by evaluation managers in agencies like yours, when commissioning and managing evaluations?

- … thoughts on the uses of AI for storage and retrieval of delivered evaluations?

APPLICA

ISMERI EUROPA

QUADRANT
CONSEIL

Panel discussion

Wrap-up