Ministerie van Buitenlandse Zaken

Welkom op het
mini-symposium van

IOB-evaluatie
Evaluatienetwerk van Vide

Articial Intelligence & Responsive Evaluation

- Steffen Bohni
- Francesco Mazzeo Rinaldi
- Tineke A. Abma
- Cristina Magro
- Frans L. Leeuw
- Peter van der Knaap

Artificial Intelligence and Evaluation.

Emerging Technologies and their Implications for Evaluation

INTEVAL & Dutch & Flemish Evaluation Societies, the Hague, NL

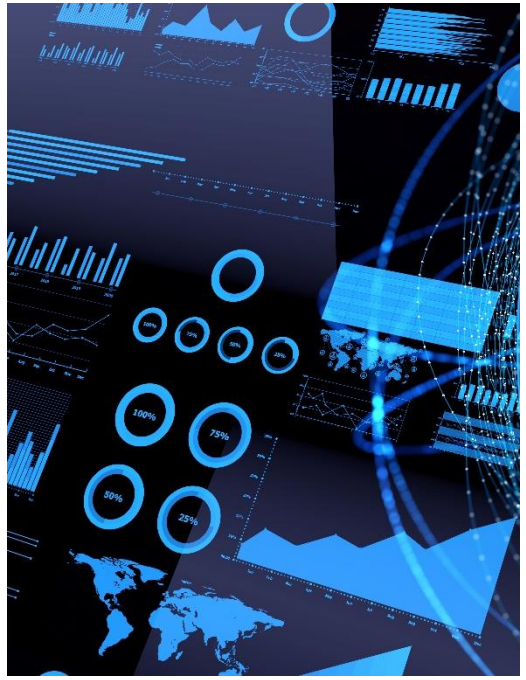
Steffen Bohni Nielsen & Francesco Mazzeo Rinaldi

28 May 2024

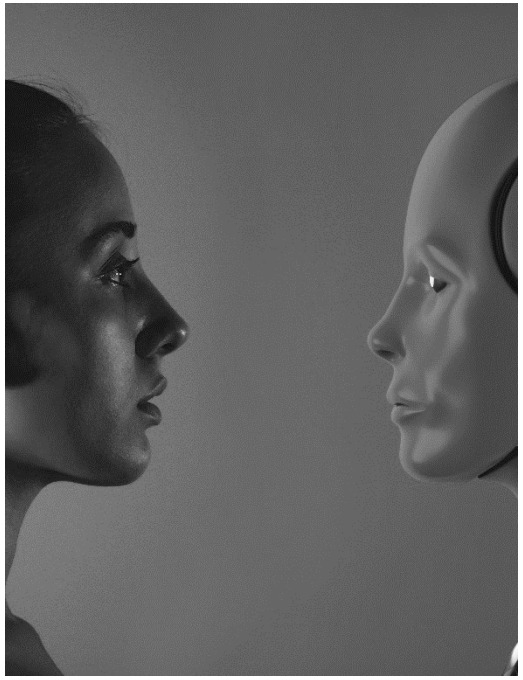
Agenda

1. About the upcoming book
2. What is AI in the context of evaluation?
3. How will evaluation practice be affected?
4. What does the future hold?

Guiding Research Questions in the Book



What are the emerging digital technologies?



What requisite skills do evaluators need?

What contribution can evaluation make to AI and vice versa?

Book - Table of contents

Section one – What is AI and Emerging Technologies

Chapter 1 : ***Evaluation in the Era of Artificial Intelligence***

Steffen Bohni Nielsen, Francesco Mazzeo Rinaldi, Gustav Petersson

Chapter 2 : ***Emerging Technology and Evaluation in International Development***

Kerry Bruce, Valentine Gandhi and Joris Vandelanotte

Chapter 3 : ***The Applications of Big Data to Strengthen Evaluation***

Pete York and Michael Bamberger

Chapter 4 : ***Ethics & Equity in Data Science for Evaluators***

Nathan Greenstein, Sung-Woo Cho

Section Three: Implications of AI

Chapter 13 : ***The Evaluation Industry and Emerging Technologies***

Steffen Bohni Nielsen

Chapter 14 : ***Artificial Intelligence: Challenges for Evaluators***

Francesco Mazzeo Rinaldi, Steffen Bohni Nielsen

Section Two – Case Applications of AI

Chapter 5: ***Extracting Meaning from Textual Data for Evaluation. – Lessons from Recent Practice at the Independent Evaluation Group of the World Bank***

Virginia Ziulu, Harsh Anuj, Ariya Hagh, Estelle Raimondo, Jos Vaessen

Chapter 6 : ***Text Mining and Machine Learning in an Evaluation of Police Handling of Cybercrime in Norway***

Tom Næss, Carolin Prabhu, Mari Mjaaland, Helge Holtermann and Lars Skage Engebretsen

Chapter 7: ***Big Data for Big Investments: Making Responsible and Effective Use of Data Science and AI in Research Councils***

Jon Holm, Denis Newman-Griffis and Gustav Jakob Petersson

Chapter 8 : ***The Use of Quantitative Text Analysis in Evaluations***

Laura Gatto and Pirmin Bundi

Chapter 9 : ***Artificial Intelligence and Text Analysis in Evaluating Complex Social Phenomena. The Russia-Ukraine Conflict***

Francesco Mazzeo Rinaldi; Elvira Celardi; Vincenzo Miracula; Antonio Picone

Chapter 10 : ***Harnessing Geospatial Approaches to Strengthen Evaluative Evidence***

Anupam Anand, Geeta Batra and Juha I. Uitto

Chapter 11 : ***The Future of Evaluation Analytics: Case Studies of Structural Causal Modeling in Action***

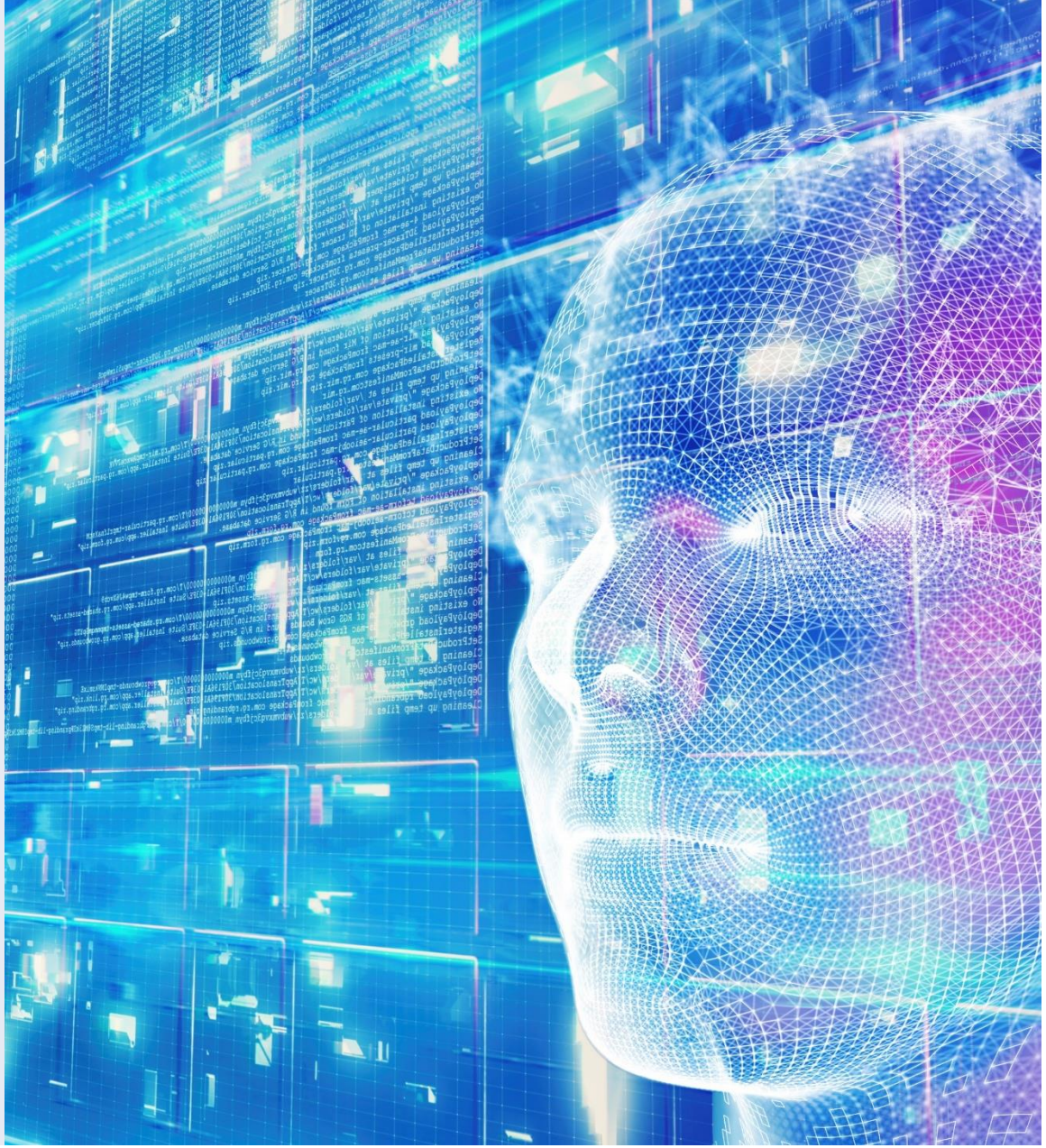
Pete York

Chapter 12 : ***The Algorithmization of Policy and Society: The Need for a Realist Evaluation Approach***

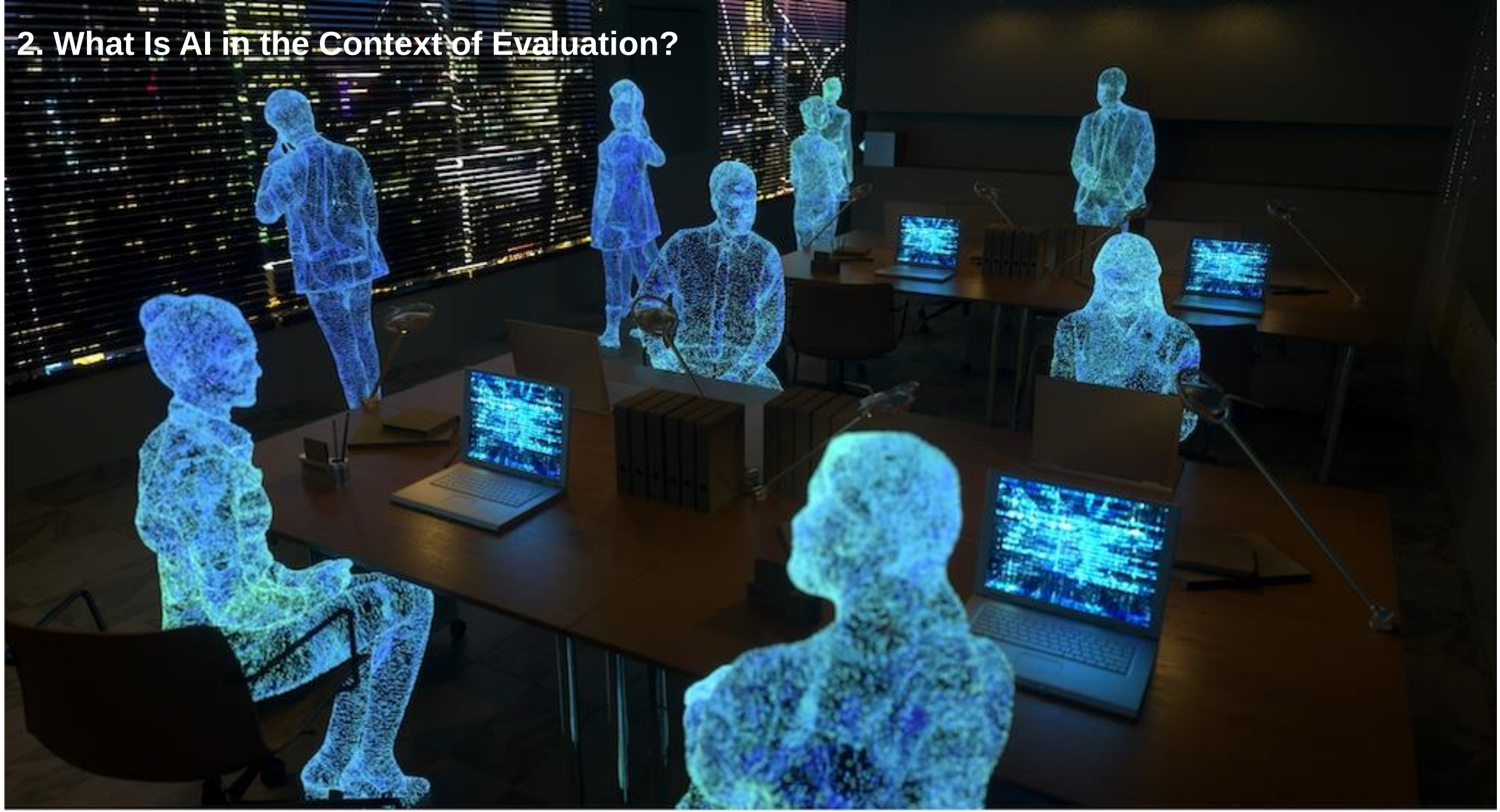
Frans L. Leeuw

Planned Publication Process

- Manuscript delivered to INTEVAl leadership end of January 2024
- Manuscript delivered to Taylor & Francins end of February 2024
- Contract signed end of March 2024
- Open Access agreement signed end of March 2024
- Book slated to be published in September 2024 ahead of EES and AEA conferences
- Conference abstracts submitted to EES and AEA



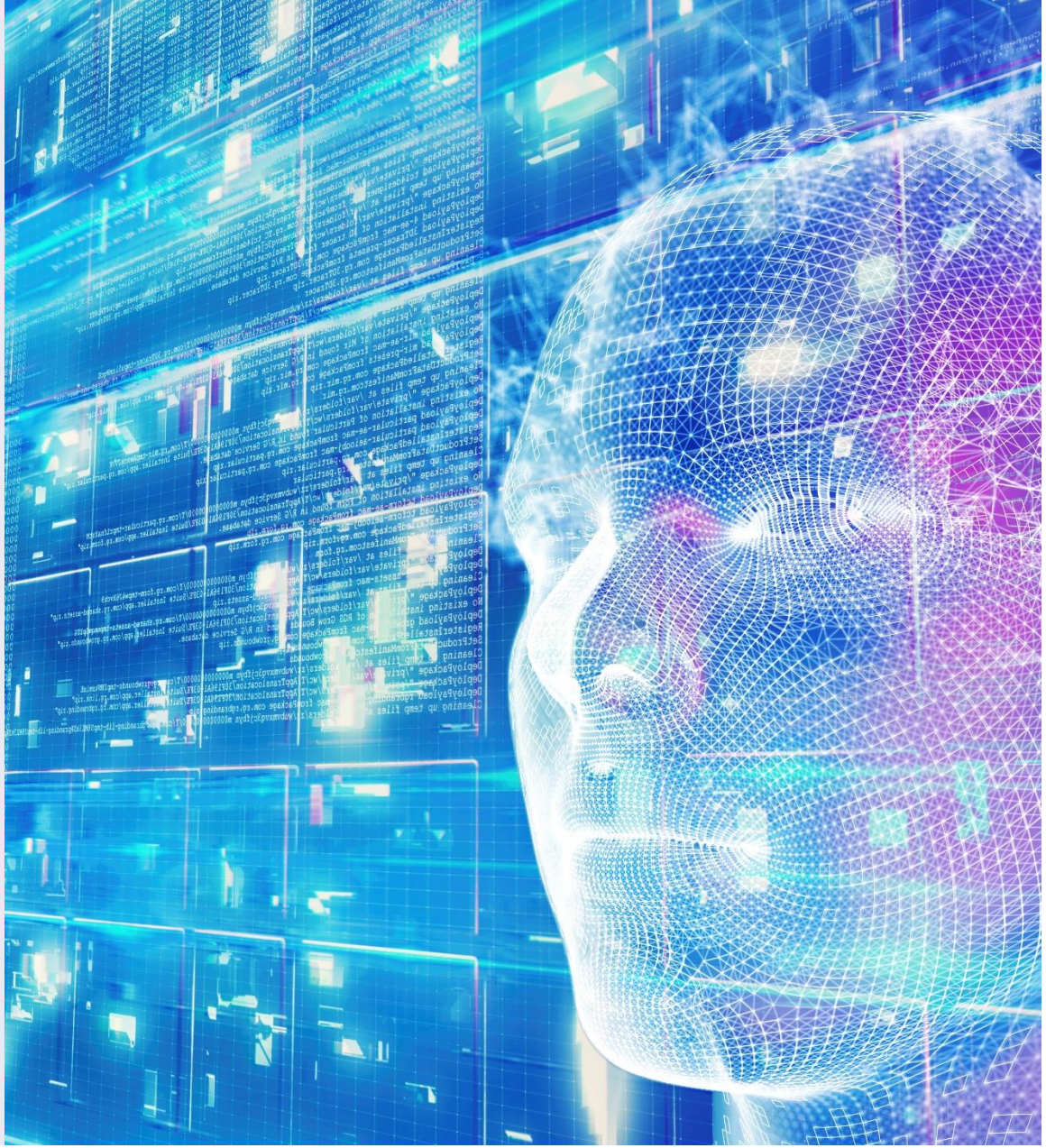
2. What Is AI in the Context of Evaluation?



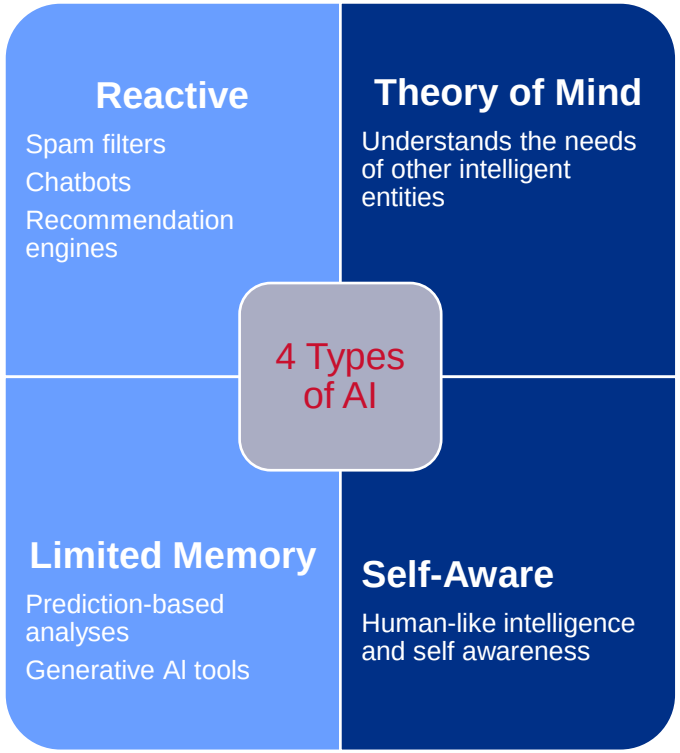
“

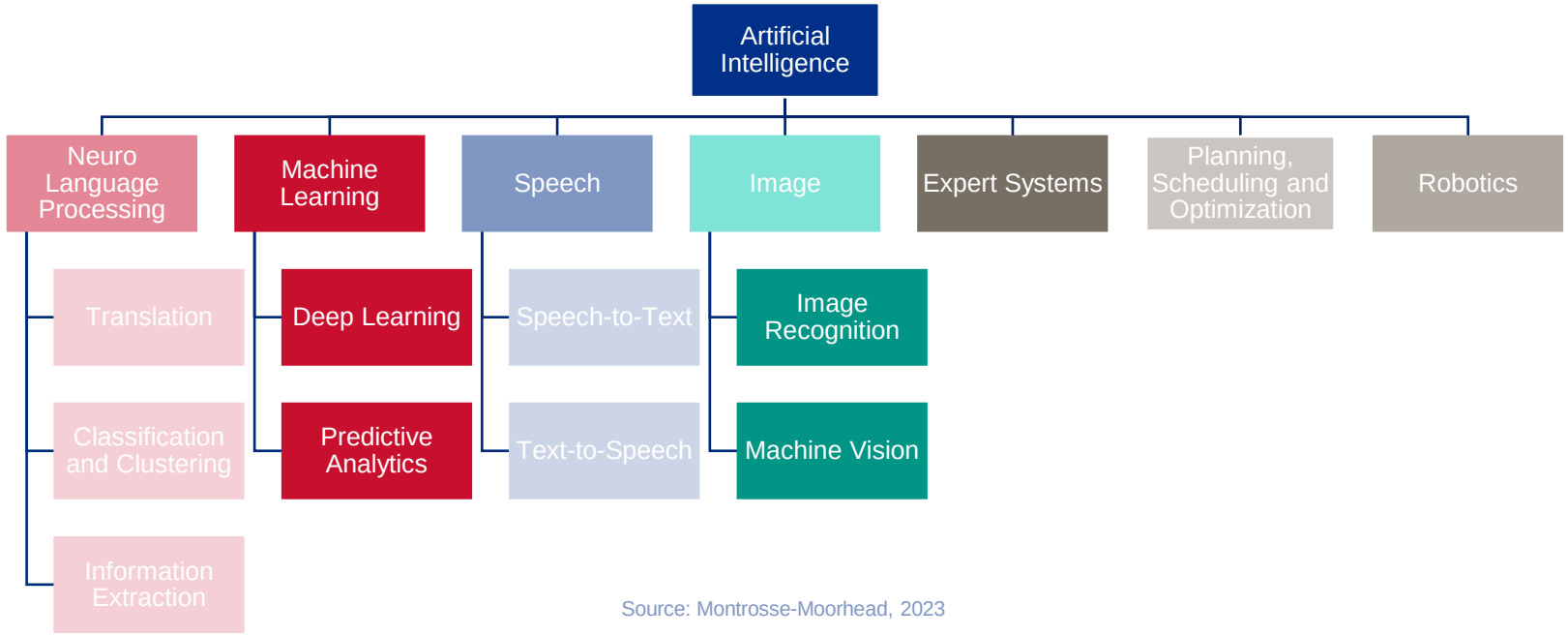
Think of Generative AI (currently) as a hyper intelligent, super creative, extremely knowledgeable, and totally unreliable assistant

Dr. Thomas Terney



Types of Artificial Intelligence





Source: Montrosse-Moorhead, 2023

Emerging Technologies available for program evaluation

(New) Kinds of data	(New) Kinds of data storage and organization	(New) Kinds of data processing

• Satellites and drones
• Social Media
• Radio call-in programs
• Internet searches
• Mobile phones
• Telecom data records
• Program administration data
• Systems data
• Large-scale survey data
• Textual data
• Internet of Things (IoT)
	• Distributed Ledger Technologies
	• Cloud Computing
	• Edge Computing
		• Machine Learning and Artificial Intelligence
		  • For quantitative analysis
		  • For text analysis
		  • For image analysis
		  • For network analysis

Source: Nielsen, 2023

AI-Evaluation

AI in the Context of Evaluation

- **Innovative methodologies**
  - Advanced Analytical Techniques
    - Predictive Analytics
    - NLP, ML, DL, GAI…
  - Complex Systems Analysis
    - Simulation Models
    - Network Analysis
  - Dynamic and Real-Time Assessment
    - Continuous Monitoring
    - Adaptive Algorithms
- **Interdisciplinary innovation**
  - Collaboration Across Fields
    - Bridging Disciplines
    - Holistic Perspectives
  - Shared Methodologies and Tools
    - Integrated Approaches
    - Cross-Disciplinary Tools
  - Innovative solution to complex problems
    - Complex Problem-Solving
    - Tailored Interventions
- **(new) challenges**
  - Data Quality and Integration
    - Data Quality
    - Integration Challenges
  - Ethical Concerns
    - Bias and Fairness
    - Transparency
  - Over-reliance on Technology
    - Loss of Human Judgment
    - Reduced Critical Thinking

3. How Will Evaluation Be Affected?

Evaluation Studies – Task level breakdown

Project Management

Design	Structuring	Data Collection	Analysis	Judging	Reporting	Utilization
• Determine evaluation question • Select research design • Select evaluation model • Determine deliverables • Stakeholder management	• Develop detailed methodology • Evaluation criteria • Evaluation standards • Indicators • Project planning • Assign tasks	• Desk study • Literature review • Collect existing data from multiple sources • Survey • Interviews • Participant observation	• Data manage-ment • Transcrip-tion • Translation • Qual coding • Quant analysis • Data and source triangulation • Performa-tive analysis	• Compare against evaluation standards • Contextual analysis • Infer judgment	• Data visualization • Findings • Conclusions Recommenda tions • Write report • Manage stakeholders	• Manage stakeholders • Planning • Adaptation of M&E system • Consultation • Advising

Monitoring and Evaluation System – Task level breakdown

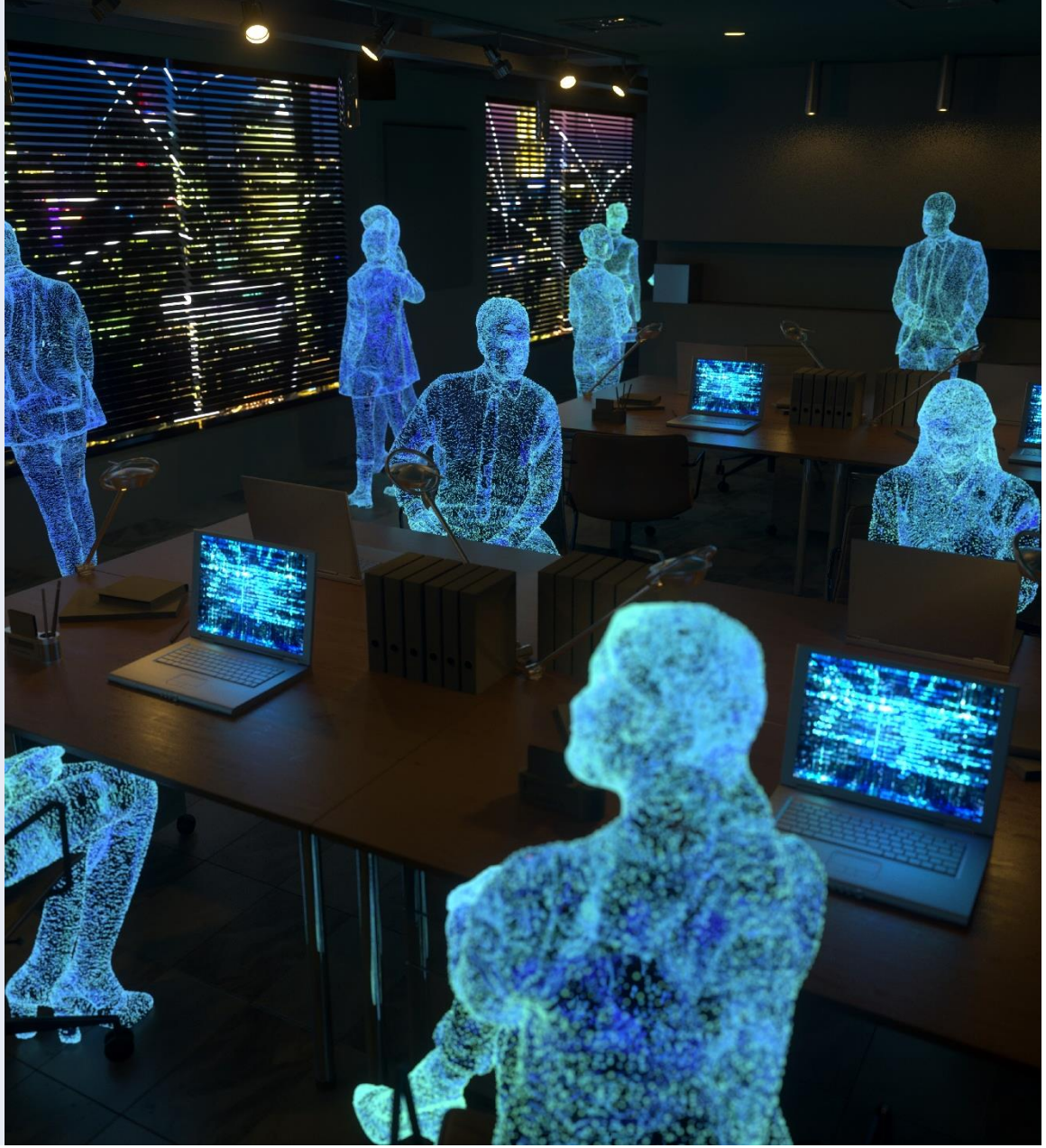
Project Management

Design	Planning	Data Collection	Analysis	Assessment	Reporting	Utilization
• M&E capacity assessment • Overall M&E strategy • Design of M&E system • Determine deliverables • Stakeholder management	• Develop indicators • 8Qs • Operationalization of all indicators • Performancestandards • Establish data mgt. function • Annual cycle planning • Assign tasks	• Desk study • Literature review • Collect existing data from multiple sources • Survey	• Data manage-ment • Quant analysis	• Compare against perfor-mance standards • Contextual analysis	• Findings • Data visualization • Tailored reporting	• Manage stakeholders • Planning • Consultation • Advising

Thematic coding of qualitative data

Using CoLoop to analyze 10 transcripts produced remarkably similar subthemes in 10–20 seconds compared to nearly a dozen hours spent coding, reviewing, and summarizing data manually.

Sabarre, Beckmann, Bhaskara & Doll, 2023:63



Coding open ended survey questions

…Avalanche produced more granular themes than those identified manually. While less frequently cited, most of these could be appropriately grouped as subthemes under our manually produced themes. There were only a few instances where Avalanche did not identify a manually generated theme…”

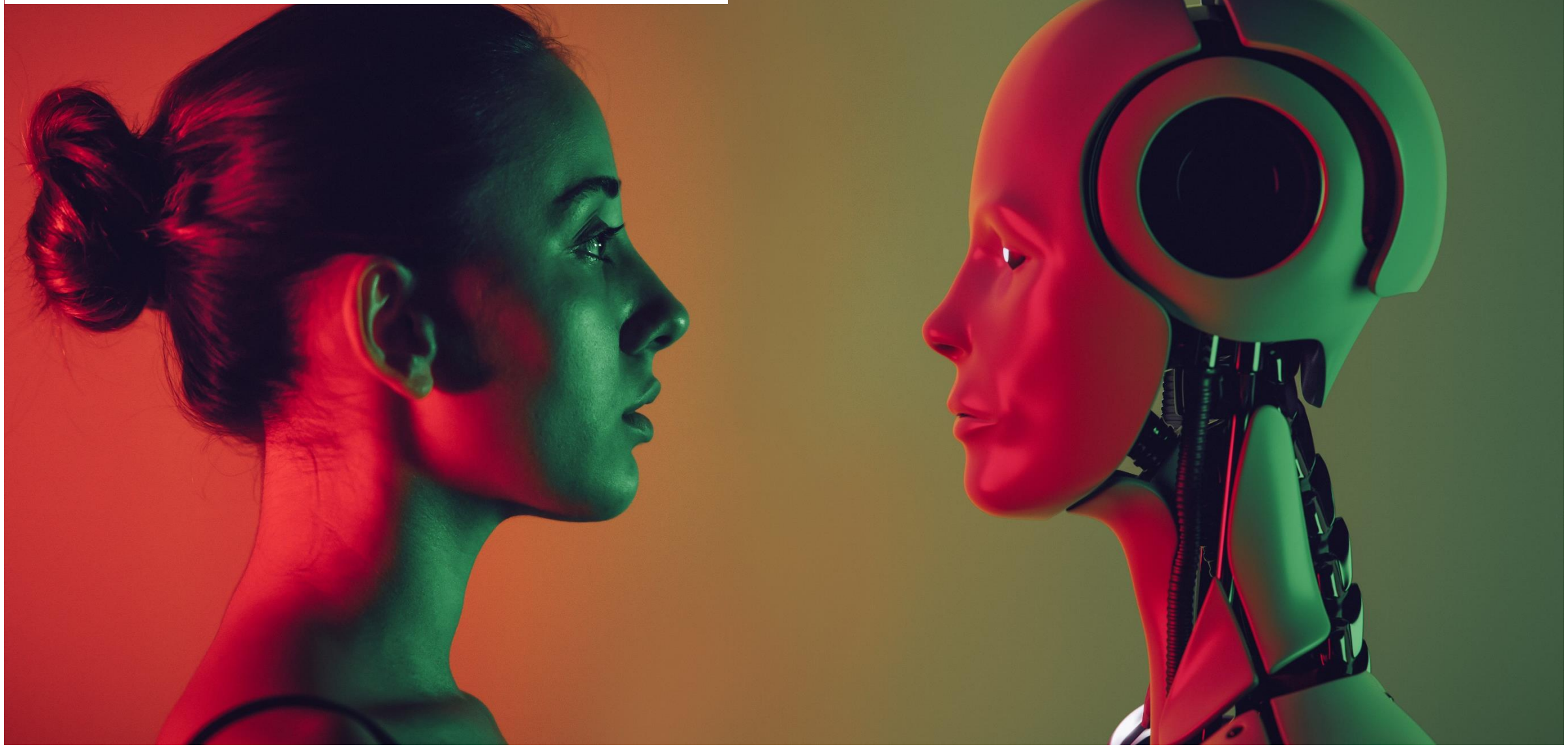
Sabarre, Beckmann, Bhaskara & Doll, 2023:65

“

Competencies that are highly social and highly creative and strategic—which may allow us to retain our specialized expertise as evaluators

Sarah Mason, 2023: 20

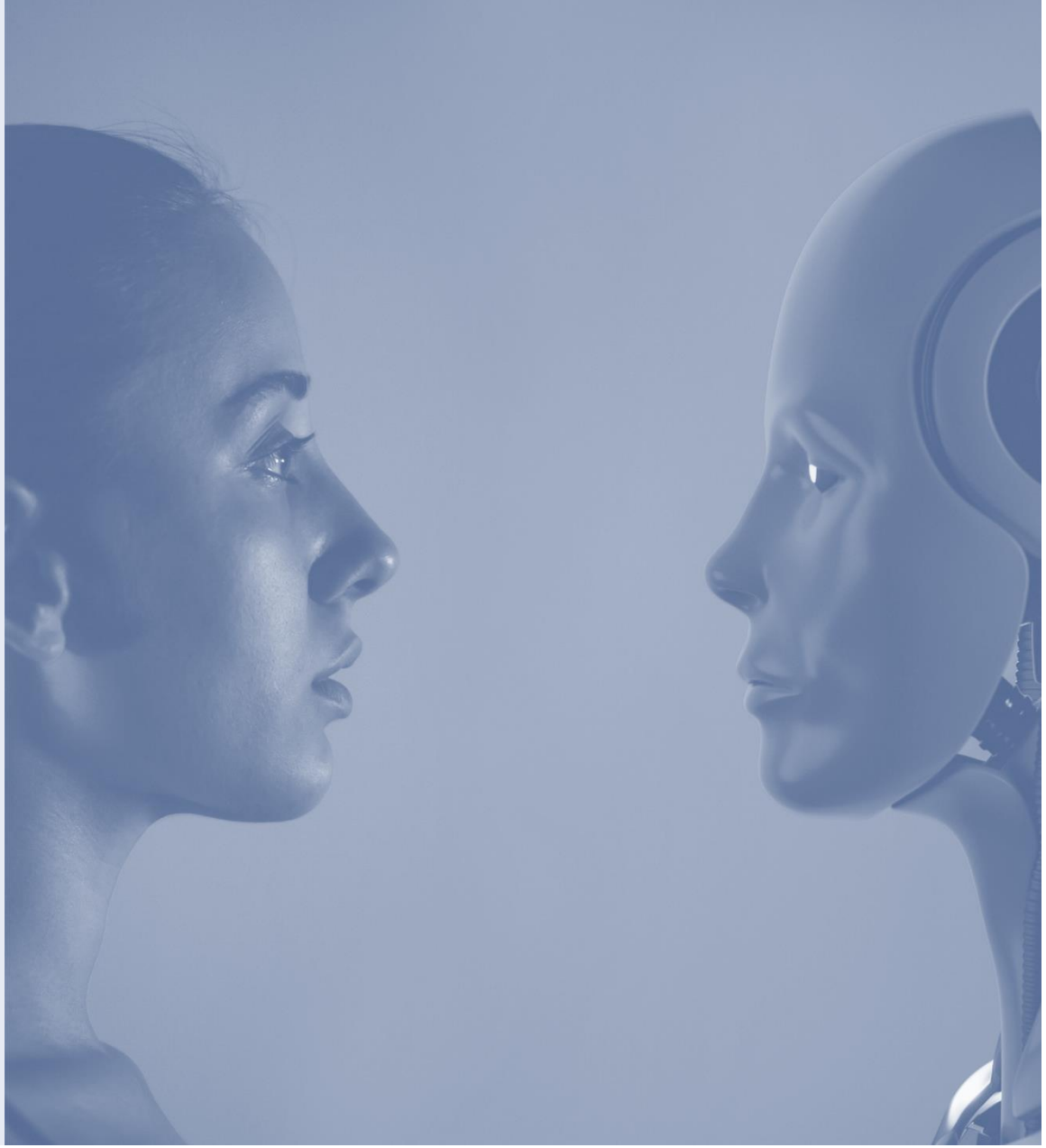
4. What Does the Future Hold?



What Evaluators Evaluate Will Change

> The evaluation community is well-positioned to provide leadership on the evaluation of and use of AI, including what criteria ought to be used.

Montrosse-Moorhead, 2023: 124



Team Expertise Composition Will Change

Subject Matter Expert

Provides in depth knowledge about the subject matter evaluated

Evaluation Expert

Provides evaluation methodology and competencies

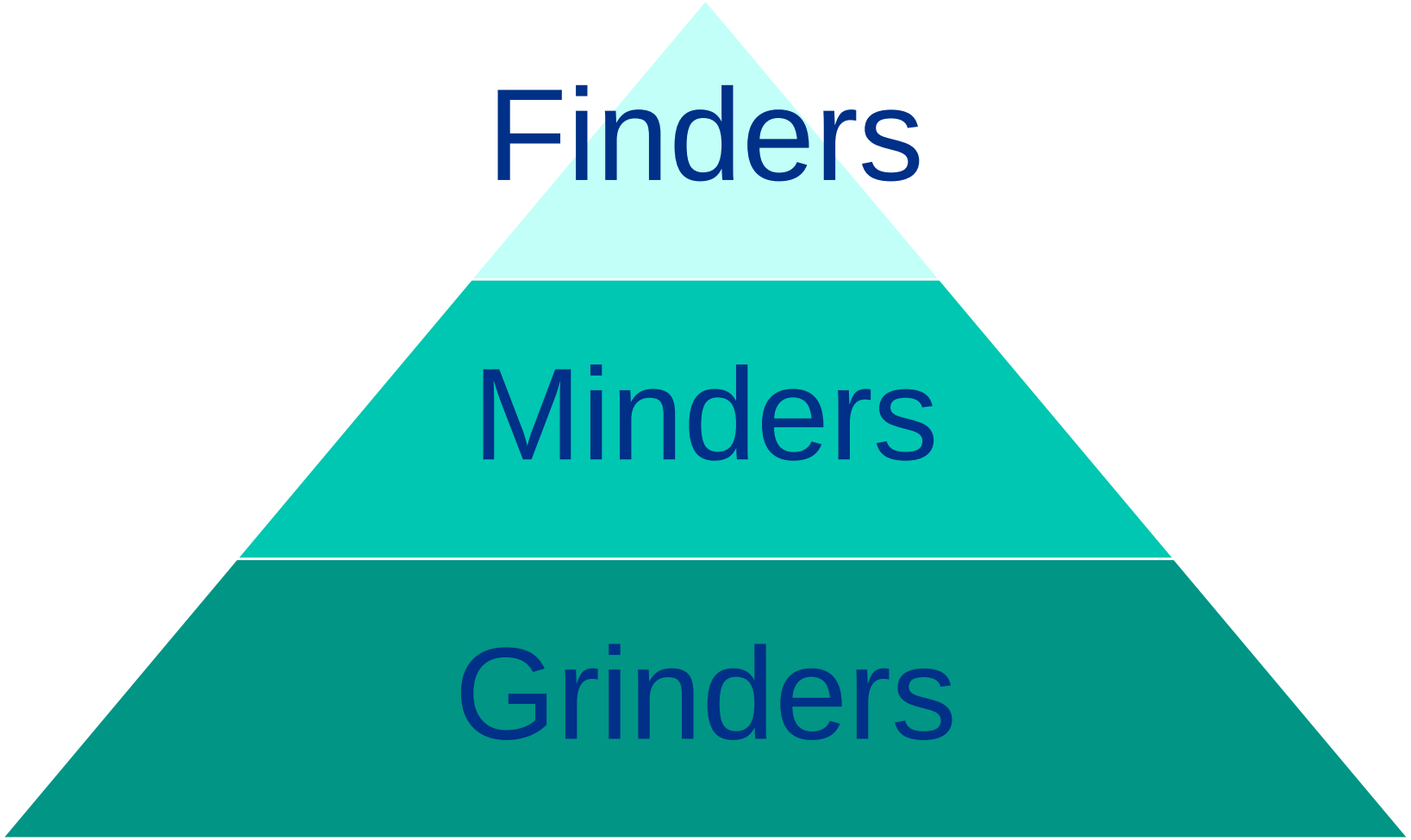
Data Scientist

Provides technical expertise in data capture, storage and processing

AI solutions

Work alongside one or several technologies

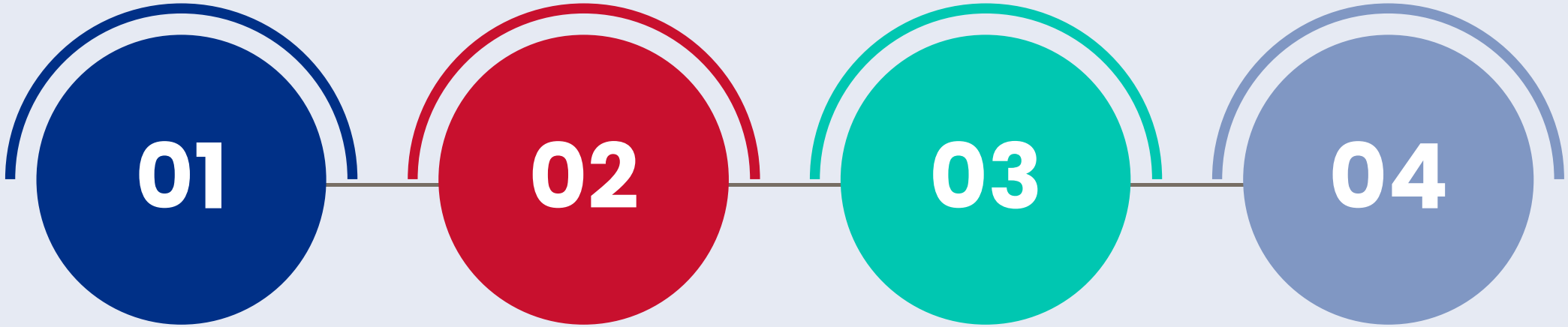
Team Experience Composition Will Change



Skills needed

1. Digital literacy
2. Data Analysis
3. Programming and coding basics
4. Data ethics and privacy
5. Collaborative skills

What Different Actors Need to Do



EVALUATORS

Upskilling, AI literacy

EVALUATION PROVIDERS

Grow talent, hire talent, collaborate to develop AI capabilities

VOPES

Evaluator competencies, upskilling programs, advocacy policy-makers

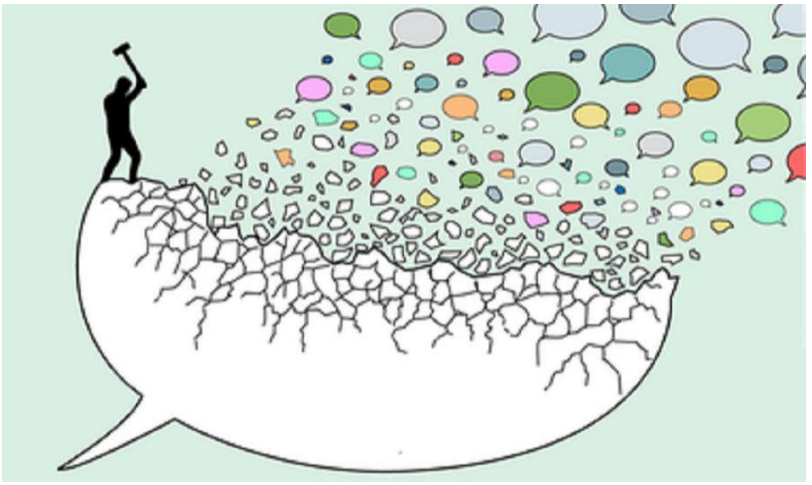
EDUCATIONAL INSTITUTIONS

Adapt education curriculum, develop competencies

Making evaluations more responsive and 'decolonized': IOB's Ambitions and Dilemmas

Ambition = more eyes and ears for stakeholders & beneficiaries' interests & needs; more collaboration with local evaluators and more local ownership & empowerment

- Quality: evaluations that take the perspectives of stakeholders & recipients seriously → What is really needed? How is it valued? Are there any unintended effects?
- Equality and reciprocity (UN treaties): why should donors have more 'evaluation rights' than partners?
- Dutch MFA → 'feminist foreign policy', reciprocity and local ownership



What are we doing?

- Country- and NGO-representatives in External Reference Groups
- Local evaluators in country studies
- Broad feed-back of results
- Joining international initiatives and promoting the country-led principle (e.g. financial and staff support for country-led Covid-evaluation Malawi)
- Capacity strengthening of the evaluation function of Southern partners

- IOB criteria and guidance → more responsive and feminist methods
- OECD/DAC Evalnet → best practices book

‘Every evaluation should be responsive.’

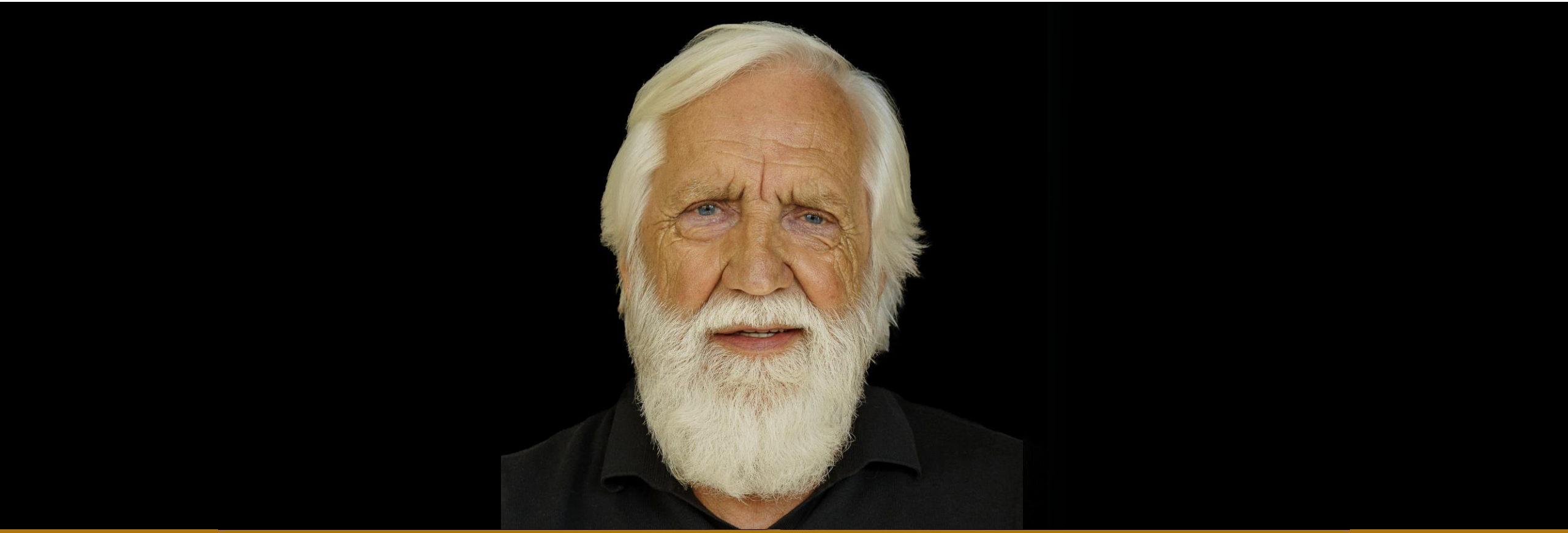
Dilemmas, paradoxes, misunderstandings

- Manageability: more local ownership of evaluation → less control over scope, research, and time?
- Focus: openness to 'their’ perspective PLUS being tied to Dutch central government evaluation questions?
- Accountability: how feasible is ‘two-way accountability’ in 2024 in Holland? → accountability to Parliament is more important than ever...

Crux = how do you find a good balance and who makes the final choices?

Leyden Academy

ON VITALITY AND AGEING



Responsive evaluation: Introduction

Prof. Dr. Tineke Abma, May 28th 2024

Prof. Dr. Tineke Abma

> Participation of older people is of great importance. It contributes to their well-being and society

- Executive-Director Leyden Academy, 2019
- Professor Older person participation LUMC, 2021

Academic career
- Professor Participation and Diversity, AmsterdamUMC, 2014-2021
- Professor Client participation elderly care, AdamUMC, 2010-2014
- Associate professor, University Maastricht, 2002 - 2009
- Assistant professor Erasmus University, 1990 - 2009

Scientific impact
- Supervised 40 PhD students
- H-index: 58, >14.000 citations

Awards
- ASPASIA laureate, NWO, 2013
- Crebolder award for impact, 2007

Shortcomings

Responsive evaluation was developed to address:

- Management-bias (Scriven) and value plurality (House)
- Under-utilization (Weiss and Patton)
- Black-box: lack of understanding in underlying mechanisms of programmes
- Democratic decision-making and having a ‘say’ by citizens (Simons)
- Under-representation of certain groups in society (Greene)

Intellectual history

Robert Stake coined the term in the ’70’s

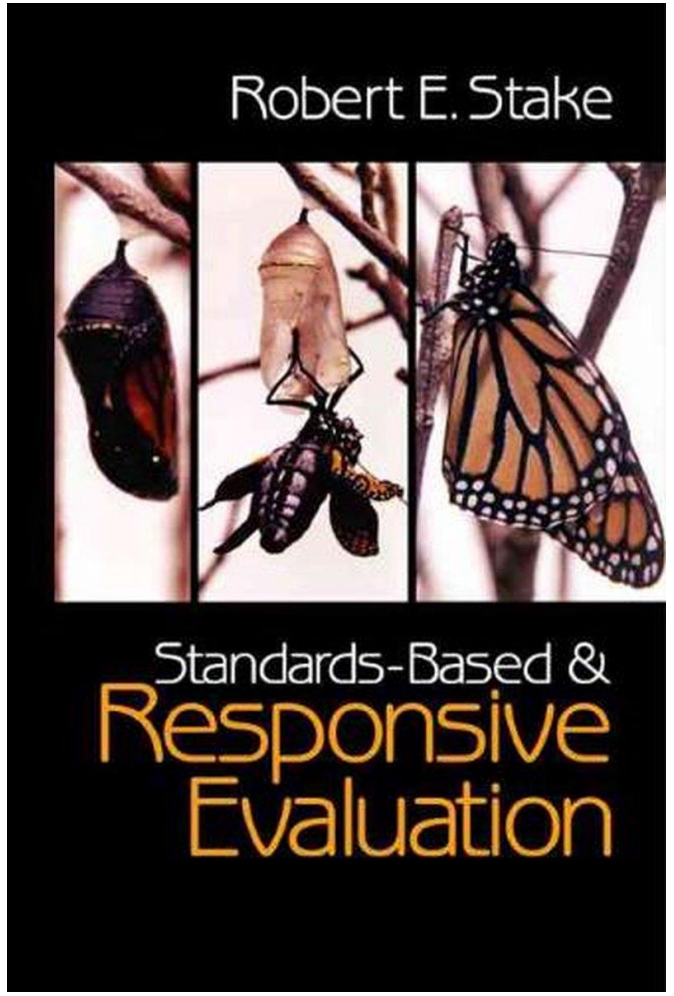
- Responsive to ‘issues’ of all stakeholders
- Actions (versus goals) and context
- Evaluator acts as social antropologist

Egon Guba and Yvonne Lincoln, 1989

- Not only ‘for’ but ‘with’ all stakeholders
- ‘Negotiation’ between stakeholders
- Evaluator as process facilitator

Comparable to:

- Democratic, participatory, transformative and empowerment evaluation and action research



Standard evaluation versus responsive evaluation

	Standard-based	Responsive evaluation
Evaluation criteria and standards	A-priori set by key stakeholder group Policy-centred Effects and outcomes	**Emergent design in dialogue with all stakeholders** **Pluralist: the values and interests of all stakeholders** **Effects, actions, implementation and context**
Evaluation process	Fixed hypothetic-deductive design Stakeholders are information-givers Monological Evaluator as expert	**Emergent design based on stakeholder issues** **Stakeholders aren partners in the process** **Interactive and dialogical** **Evaluator as facilitator**
Learning	Individual/cognitive phenomenon Begins after the evaluation data More knowledge 'about' a programme	**Collective/social phenomenon** **Begins during the process of evaluation** **Personal and mutual understanding among stakeholders for collective action**

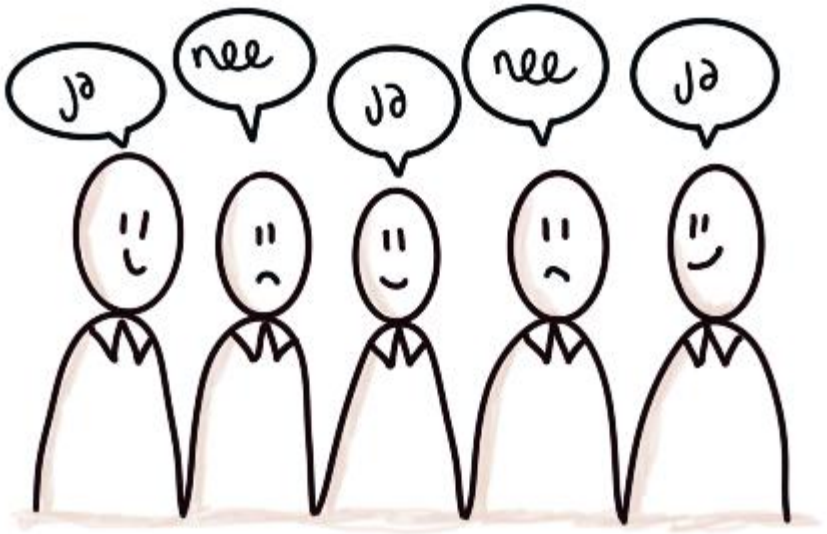
Responsive evaluation

Goals

- Insights in the various stakeholder perspectives and underlying values
- Personal and mutual understanding through dialogue
- Mutual learning, change and empowerment: collective action to improve practice

Multi-stakeholder process

- To deal with power differentials multi-phased process
- Start with the stakeholder group least heard; their issues are often unknown
- Deliberation and dialogue with stakeholder groups (homogeneous) and later between groups (heterogeneous)



Multi-phased and cyclic process

1. Exploration: identification stakeholders, creating social conditions for participation
2. Consultation: all stakeholders are consulted to identify their issues and concerns
3. Collaboration: validation and deepening of the issues in homogeneous groups
4. Dialogue: heterogeneous exchange of issues and perspectives
5. Integration: formulating an agenda for negotiation and plan for action to improve practice
6. Implementation: Action plans are carried out and monitored

Design and methods

- Broad range of methods, qualitative and quantitative

<u>Steps and considerations</u>

1. Consult and make a choice in collaboration with stakeholders, including citizens
2. Which methods are appropriate to foster maximum participation and interaction?
3. Which methods are suitable to foster personal and mutual understanding?

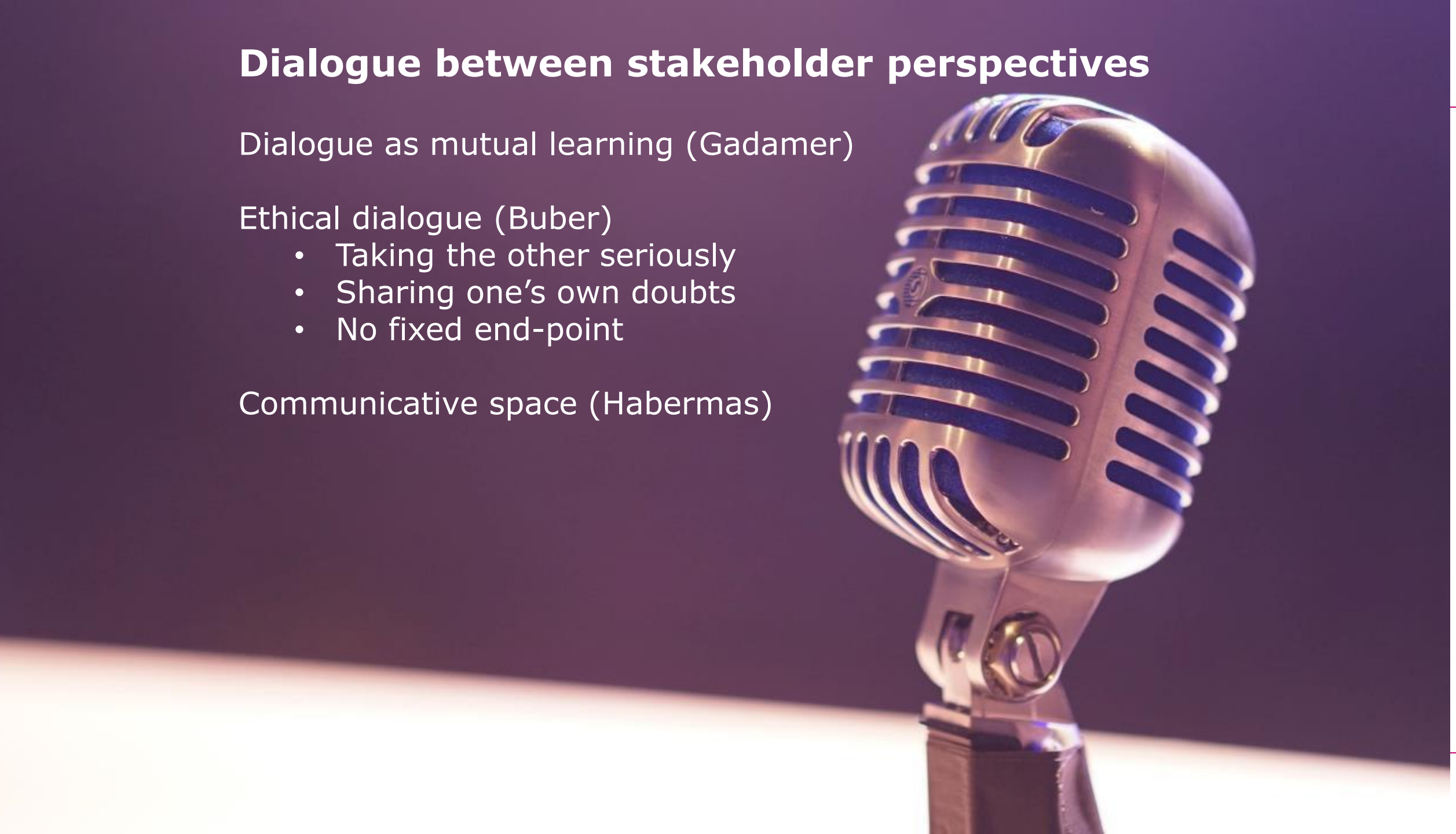
Dialogue between stakeholder perspectives

Dialogue as mutual learning (Gadamer)

Ethical dialogue (Buber)

- Taking the other seriously
- Sharing one’s own doubts
- No fixed end-point

Communicative space (Habermas)



Conditions

Willingness among all stakeholders

- To actively participate in the process
- To share power with other stakeholder groups
- To change and transform in dialogues

Epistemology and quality criteria

Social constructivism

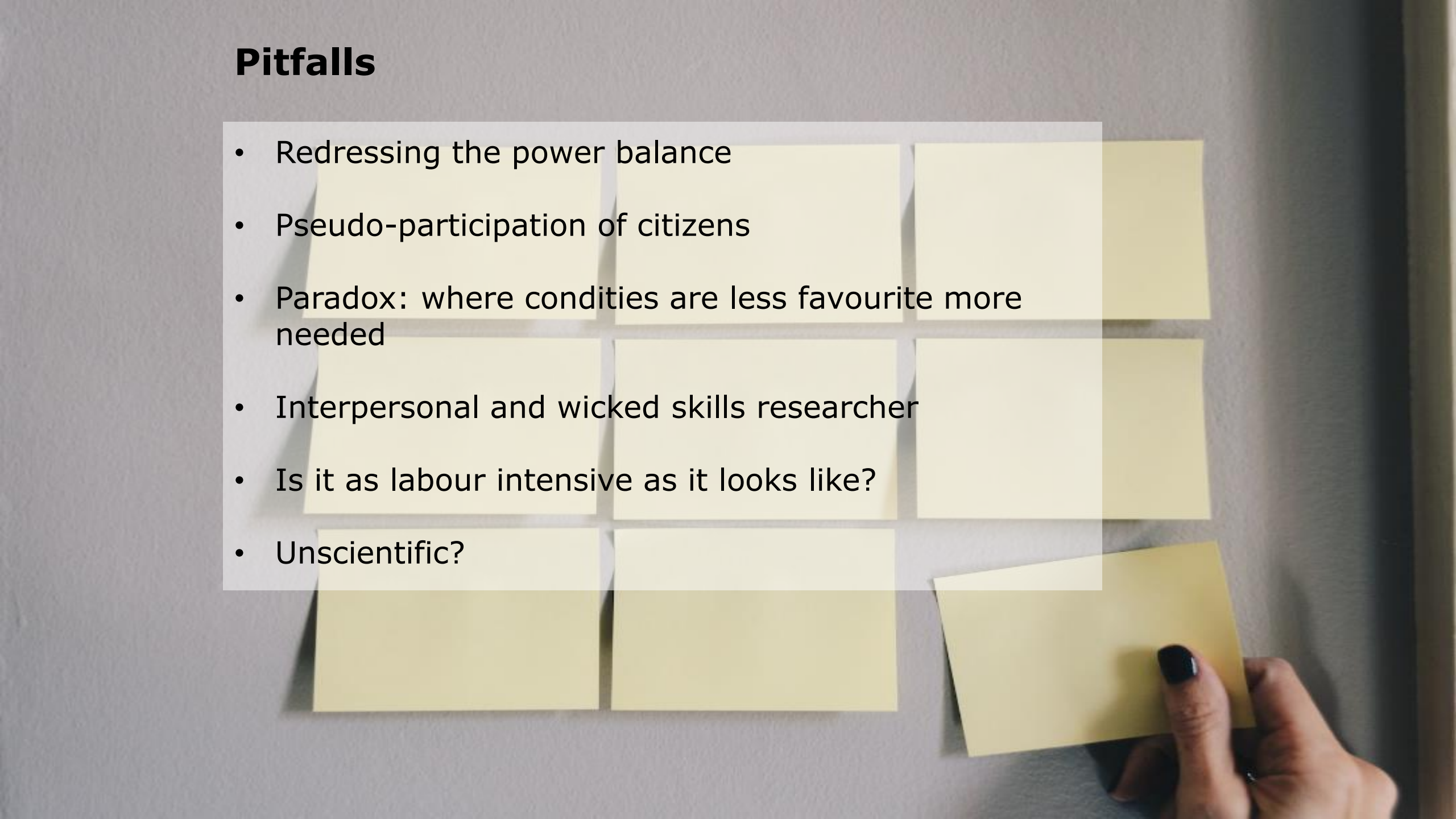
- Human beings are constructing meaning in interactions
- Researcher and researched are mutually influencing each other > ‘flattening’ hierarchies: expert and lay-persons
- Knowledge is context and time-bound
- Values are influencing the evaluation

Quality criteria

- Credibility criteria
- Fairness and epistemic justice
- Authenticity criteria
- Hermeneutic dialectic process

Pitfalls

- Redressing the power balance
- Pseudo-participation of citizens
- Paradox: where condities are less favourite more needed
- Interpersonal and wicked skills researcher
- Is it as labour intensive as it looks like?
- Unscientific?



Added value and potential

- Feedback from citizens, including unintended side-effects
- Insight in complex issues from multiple stakeholder perspectives
- Impact during the evaluation process
- Being heard and co-ownership fosters dissemination and implementation

Summary

- Responsive evaluation offers a vision and rationale for evaluation
- Central is the dialogue between stakeholders
- Multi-stakeholder, multi-phased process
- Active involvement of all stakeholders in the process
- Useful for policy fields characterized by ambiguity and diverging perspectives

Read more?

- Abma, Tineke; Groot, Barbara (2023) Participatory Governance of Healthcare: Centring Dialogue and Interrupting Epistemic Injustice. Journal of Dialogue Studies . 2023 Special Issue, Vol. 11, p118-139.
- Abma, T. A. (2020). Ethics work for good participatory action research. Beleidsonderzoek Online. doi:10.5553/BO/221335502020000006001
- Abma, T. A., Visse, M., Hanberger, A., Simons, H., & Greene, J. C. (2020). Enriching evaluation practice through care ethics. Evaluation, 26(2), 131-146.
- Abma, TA, Cook, T, Rämgård, M, Kleba, E, Harris, J & Wallerstein, N (2017) 'Social impact of participatory health research: collaborative non-linear processes of knowledge mobilization' Educational Action Research, vol 25, no. 4, pp. 489-505.
- Abma, TA, Leyerzapf, H & Landeweer, E (2017) 'Responsive Evaluation in the Interference Zone Between System and Lifeworld' American Journal of Evaluation, vol 38, no. 4, pp. 507-520.
- Abma, T.A. en G.A.M. Widdershoven (2006) Responsieve methodologie. Interactief onderzoek in de praktijk, Den Haag: Lemma.
- Abma, T.A. (2005) The practice and politics of responsive evaluation, The American Journal of Evaluation, 27(1): 31-43.
- Abma, T.A. (2004) Responsive evaluation: Its meaning and special contribution for Public Administration, Public Administration, 82(4): 993-1012.
- Abma, T. A., & Noordegraaf, M. (2003). Public managers amidst ambiguity: Towards a typology of evaluative practices in public management. Evaluation, 9(3), 285-306.

Leyden Academy

ON VITALITY AND AGEING

Rijnsburgerweg 10 | 2333 AA Leiden | 071-524 0960 | info@leydenacademy.nl | www.leydenacademy.nl

Unpacking the single-story experience for a healthier planet

Cristina Magro, Brazil

Independent Evaluator, IEAc, INTEVAL

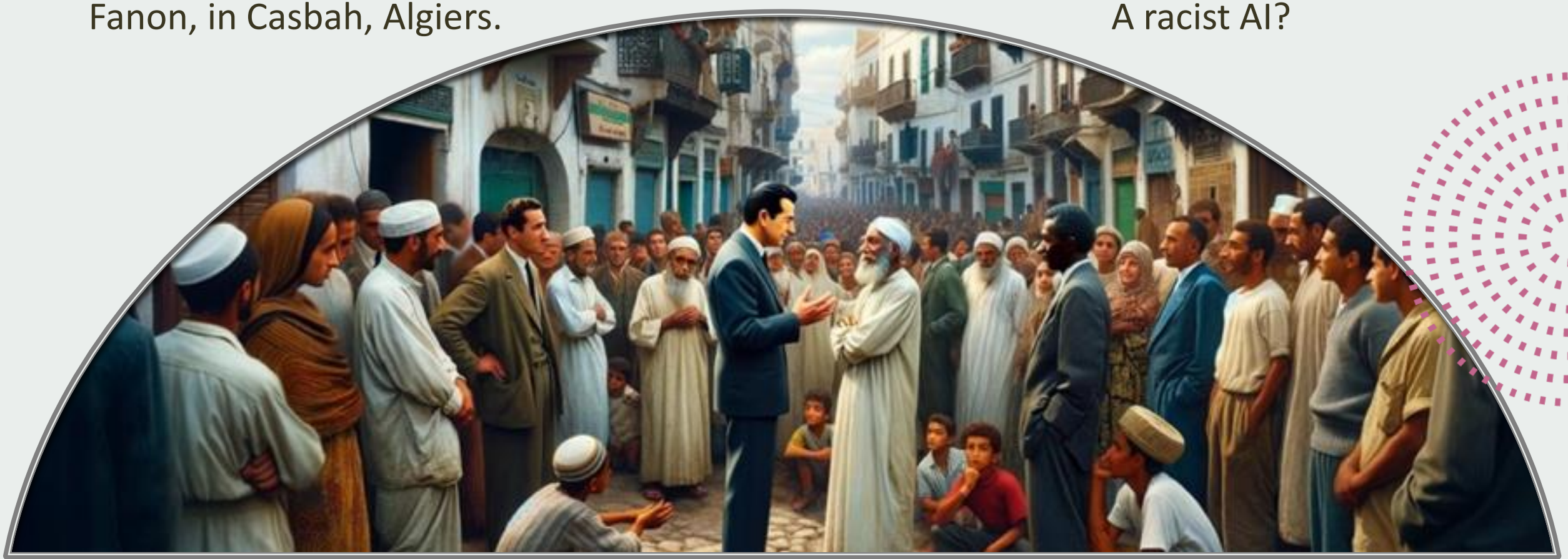
28 of May 2024

IOB evaluation, Ministry of Foreign Affairs, The Hague

Frantz Fanon, Dr. White

The AI system DALL·E 2 was prompted to generate images of the Black Caribbean psychiatrist, thinker, writer, anti-racist and decolonization activist, Dr. Frantz Omar Fanon, in Casbah, Algiers.

A racist AI?



A Black woman in a slum setting

The State Representative Renata Souza is graduated at Fed. Un. of Rio de Janeiro, and is the president of the Facial Recognition Committee in Rio de Janeiro Parliament. She prompted AI to create her image with a Disney inspiration. She pictured herself as a black woman, with curly hair, African print clothes, in a slum setting – where she was raised and lives. Stereotyped AI?

Source: Deivison (Nkosi) Faustino and Walter Lippold. *Frantz Fanon, o "Dr. Branco" e o colonialismo digital. Retrieved from* https://outraspalavras.net/outrasmidias/frantz-fanon-o-dr-branco-e-o-colonialismo-digital/ on 04/04/2024.

Setting the stage

- Diversity, equity, inclusion, social justice
- Culturally Responsive, Culturally Responsive Indigenous, Culturally Responsive and Equitable: supported by an optimistic, bold and inspirational literature and a lot of effort to develop and refine them.
- Decolonization studies
- Responses to current sources of imbalance named as
  - *colonization*, or *the single-story experience*, or **the *paradigm of a UNIverse*, that rules out alternative thinkings, ways of living and varied cosmogonies, and don't acknowledge the meaningful influence that all these, together with language, have on behavior, perception and cognition.**

Answering questions

How to address the sensitive situation of considering indigenous ways of living, thinking and being in evaluation and policy making?

Q1:

Could the reflection about indigenous knowledge in evaluation lead to better results in efforts to *decolonize* evaluation research in development (or: international) cooperation?

Indigenous evaluation and decolonization

Acknowledging indigenous knowledges in evaluation ***is one necessary aspect of decolonization and these should not be conflated:***

- Indigenous communities may disclose knowledge systems and issues which are incomparable to others we are engaged with in the decolonization efforts.
- Indigenous communities have peculiarities that make evaluation more sensitive and prone to imputations of western categories of intellectual developments (e.g. systems thinking). *Epistemological awareness is necessary*
- But yes, *a critical reflection* about indigenous knowledge in evaluation could support the efforts to *decolonize* evaluation **and vice versa**, ***provided that indigenous epistemologies, ways of living, cosmogonies, and realms of existence be acknowledged, taken as legitimate and truthful, and engaged in casting the worlds they live in*.**

Answering questions

How to address the sensitive situation of considering indigenous ways of living, thinking and being in evaluation and policy making?

Q2:

Could the concept of *positionality* in the context of guaranteeing the rights of the indigenous populations help?

Positionality and evaluation in the indigenous context

- Positionality theory underscores the *situatedness* and *contingenciality* of identity, which is "determined by where one stands in relation to 'the other'" (Merriam et al. 2001, p. 411) at a specific moment in time and place.
- The concept of *positionality* is the current flag indigenous populations are raising to warn us that the era of considering them infants, disfavored, uncapable, marginal, speechless, should be over.
- Indigenous peoples, including indigenous evaluators, *need to be recognized as fully capable of designing interventions necessary to support their livelihoods, and to follow up and learn implementation, even if **we** don't really understand their methods and processes.*

Positionality:
a cross-cutting and integrating concept

- It is crucial to realize that ***everyone has his/her positionality(ies),*** as we are speaking of a social, historical and cultural localization in the social chess board.
- This awareness is paramount to decolonize knowledge and refute any epistemological neutrality and singularity.

Answering questions

How to address the sensitive situation of considering indigenous ways of living, thinking and being in evaluation and policy making?

Q3:

What are the practical and epistemological implications of all this?

Evaluation in indigenous contexts and beyond: epistemological implications

A. Foundational conditions:

- To acknowledge multiple epistemologies and multiple realities as legitimate and yielded by the typical human lives's constitution in eco-social communities, coordinating actions in language and culture along history.
  - Corollary: to acknowledge the impossibility of referring to an independent reality, out of the culturally situated abstractions promoted by the western epistemology.
- To avoid the imputation of categories of reasoning and theories to indigenous ways of thinking.

Evaluation in indigenous contexts and beyond: epistemological implications

B. Implications:

1. To allow indigenous communities propose and guide the support wanted.
2. To quit the engrained Western separability between observer and phenomena observed, system and context.
3. To welcome the cyclic relation between practice and theory (as a reformulation of experience), moving away from traditional dependence on methodology.
4. To nourish (self-)awareness.

Many thanks!

Cristina Magro

magro.cristina@gmail.com

Ministerie van Buitenlandse Zaken

Thank you all for coming to The Hague!