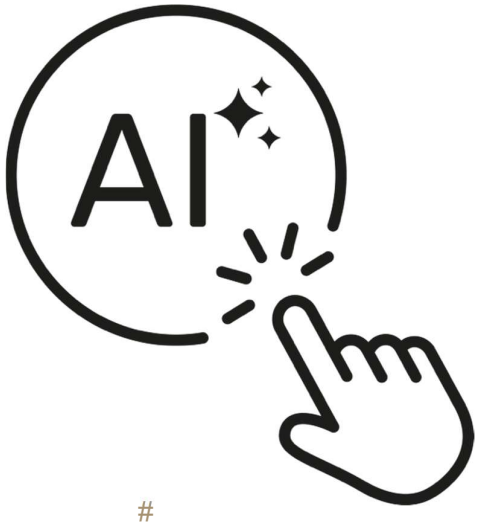


#

MRS Report
Advanced Insights & Analytics Council (AIA)
March 2026 (v2)

By Sabrina Trinquetel

Systemic Inclusion by Design

An inclusive-first framework to strengthen AI quality and integrity in research

80 years of MRS
Evidence matters
more than ever

Contents

Executive Summary	**3**
Section 1 Introduction	**5**
Section 2: Ethical, Responsible and Inclusive AI	**6**
Section 3: Inclusive AI as a Strategic EDI Lever	**7**
Section 4: What does this mean for Research?	**8**
Section 5: Existing AI Ethics and Fairness Frameworks: From Macro to Operational	**10**
Section 6: Systemic Inclusion by Design: A Framework	**16**
Section 7: Inclusion Tools in more detail	**19**
Section 8: Stage 1 Tool 1: Systemic Inclusion Canvas	**20**
Section 9: Stage 3 Tool 2: Bias Audit Protocols	**24**
Section 10: Stage 4 Tool 3: Model Scoring Cards	**27**
Conclusion	**29**
Appendix	**31**

Executive Summary

Artificial Intelligence is now integral to the future of the research sector, from data collection to insight generation. But AI is systemically biased[1]. As such it risks creating inaccurate insights and driving increased societal inequity.

As adoption accelerates, we must ensure we approach these new tools with the same rigor and scrutiny applied to all new research methods to ensure they are as accurate and high quality as they can be.

By taking an inclusive-first approach to AI, we aim to address the impact on accuracy existing biases create and strengthen the integrity and precision of our tools. Inclusion can become not just a moral ideal, but a driver of methodological excellence.

A compelling byproduct, inclusive-first AI can also be the conduit to rebuild the positive engagement of EDI practises across the wider organisation. Through such high-profile integration inclusion can take centre-stage and become harder to exclude in the future.

This paper invites us to consider a systemic inclusive approach to AI, built into every decision we make via a comprehensive framework designed around the existing stages of AI development from problem scoping through to model training, deployment and governance. Alongside this we consider aspects such as the equity of voices, power dynamics, algorithmic fairness, consent and many other factors.

As a practical starting point, this paper contains **three bespoke tools** that can be used immediately by practitioners at different stages of development to activate this systemic inclusive shift:

1. **Systemic Inclusion Canvas**: Stage 1 - Problem Definition and Scoping
2. **Bias Audit Protocols**: Stage 3 - Model Development and Training
3. **Inclusivity Scoring Cards**: Stage 4 - Deployment

Bias is not new, but AI tools have the worrying capability to amplify these biases beyond anything seen before.

[1] https://dl.acm.org/doi/10.1145/3457607

This is a practical guide for practitioners to consider integrating an inclusive-first approach to AI design to increase accuracy and decrease inequity across our sector.

“ So you see, our technology is not some grand mysterious thing outside of our control and even our comprehension. Our technology, at its most basic level, is a reflection of our humanity”

Olivia Gambelin,
Founder of Ethical Intelligence[2]

[2] Responsible AI: Implement an Ethical Approach in Your Organization (2024) Olivia Gambelin

SECTION 1

Introduction

The research sector is undergoing rapid transformation as AI becomes embedded in how we collect, analyse, and interpret data. But while the technology is moving fast, the ethical safeguards around it haven't kept pace, particularly when it comes to fairness, representation, and inclusion.

There are a number of ethical considerations that come into play when integrating AI into our work including transparency, accountability, privacy and data protection, consent, fairness, manipulation and social responsibility. These are covered in more detail within the MRS AI Guidance found here: https://www.mrs.org.uk/pdf/AI+RelatedTechnologies_MRSGuidance.pdf

One particular ethical concern is the inherent bias that is present in all aspects of research, and that now with AI risks not only being overlooked but amplified creating bigger equity gaps than before.

We believe that practitioners would benefit from having more tools to address bias in AI development and therefore have worked to design core materials that will support all organisations and practitioners to positively integrate and start this journey.

The risk is great; without intentional inclusive design, practitioners could replicate the same structural inequities our insights are helping to address.

To address this requires a shift: from passive to proactive, ethical shaping of AI systems that reflect the full diversity of the people they aim to understand. This includes fairness checks, inclusive sampling approaches, transparency standards, and a shared commitment to equity across the AI development lifecycle, all central tenets of **Systemic Inclusion by Design**.

The stakes are high. Poorly designed models don't just lead to flawed insights, they damage credibility, erode trust, and risk real-world harm. In a time when climate, AI, and inequality are converging as defining forces, the research sector must step up not only to ensure our tools are accurate, but to ensure they are just.

This paper lays the groundwork for a practical, industry-specific **Inclusive AI Framework**, a first step in helping practitioners, agencies and research leaders embed ethical design and inclusive thinking into every AI-enabled insight we produce.

SECTION 2:

Ethical, Responsible and Inclusive AI

There are a number of existing terms used in the field of Inclusive AI, which have been described below. For the purposes of this paper, we will be referring to Inclusive AI as an umbrella concept that aims to support the three angles below. However, there are vast resources that sit behind each of these elements that can be explored further by the reader.

- **Ethical AI** refers to the principles that guide AI systems to do no harm ensuring fairness, accountability, transparency, privacy, and human oversight. It's about aligning technology with moral values and societal expectations.

- **Responsible AI** builds on ethics by focusing on how those principles are operationalised. It includes the governance, risk management, and processes that ensure AI systems are safe, trustworthy, and legally compliant from design to deployment and beyond.

- **Inclusive AI** goes further, asking not just 'Is this system fair, but who is it fair for?' It emphasises the proactive inclusion of diverse voices, experiences, and identities particularly those historically marginalised in both the data and design process. Inclusive AI demands that we recognise intersectionality, address systemic bias, and ensure that AI works for everyone, not just the majority.

Together, these concepts form a foundation of AI that is not only powerful, but principled and essential for a research sector that prides itself on accuracy, representation, and public trust.

SECTION 3:

Inclusive AI as a Strategic EDI Lever

In the shadow of EDI program rollbacks, reassigned priorities and budget constraints across industries, EDI commitments are plateauing. External pressures such as economic uncertainty, societal pressure, evolving regulation and shrinking resources are making it harder to accelerate meaningful change.

"D&I strategies are becoming increasingly insular, initiatives are being shelved to prioritise other business goals."[3]

Adopting an inclusive-first AI approach offers a unique opportunity to reverse this trend, bringing EDI out of the shadows and back to the forefront of organisational attention.

The *AI for Good Impact Report* (2024)[4] emphasises that inclusive, representative AI frameworks are not only ethical imperatives but key drivers of business value and trust.

Leveraging AI in this way allows organisations to rebrand EDI as essential rather than peripheral. Bringing EDI experts into conversations with AI engineers, means that the focus is not just risk mitigation but driving innovation.

The very act of designing AI to be inclusive revitalises existing EDI frameworks that may have been deprioritised. Organisations committed to inclusive AI are, in effect, forced to confront and operationalise their equity, diversity, and inclusion commitments, giving EDI visibility, momentum, and strategic weight without it being treated as an afterthought.

It also becomes structurally durable, baked into the system, difficult to remove or dismiss in the future when budgets are scarce, positioning equity as a core driver of innovation and organisational success.

[3] The Tech Talent Charter's *Diversity in Tech 2024* report techtalentcharter.co.uk

[4] https://s41721.pcdn.co/wp-content/uploads/2024/10/AI-for-Good-Impact-Report.pdf

SECTION 4:

What does this mean for Research?

The core of research is accurate, reliable, and representative insight. When AI tools are embedded into this process, they hold incredible potential to enhance scale, speed and depth of understanding. But this potential comes with great responsibility.

Without **Systemic Inclusion by Design**, AI-driven research risks perpetuating and amplifying the same social inequities and biases that already exist:

1. **Example: Gender and LGBTQ+ Representation**

   Research tools trained on biased data may misgender participants or fail to capture the diversity of LGBTQ+ identities. GLAAD[5] highlighted how AI systems often erase or misrepresent LGBTQ+ people, reinforcing exclusion and invisibility in insights that organisations rely on to make product and policy decisions.

2. **Example: Ethnic and Cultural Bias**

   AI language models and sentiment analysis tools often perform worse on non-majority dialects or cultural references. This leads to underrepresentation of minority voices or misinterpretation of responses, skewing data and disadvantaging these communities. For example, African American Vernacular English (AAVE)[6] has historically been poorly understood by many AI systems, leading to misclassification or exclusion.

3. **Example: Socioeconomic and Geographic Bias**

   Data collection methods relying heavily on digital or mobile platforms risk excluding lower-income or rural populations with limited internet access[7]. This can falsely inflate the perceived preferences or needs of more connected urban populations, resulting in products or services that do not meet broader societal needs.

[5] https://glaad.org/smsi/2024/focus-on-ai/
[6] https://arxiv.org/abs/2410.11005
[7] https://www.un.org/en/delegate/itu-29-billion-people-still-offline

These biases have real-world consequences:

- They undermine trust in research insights, as decisions based on incomplete or skewed data fail to resonate with or serve large parts of the population.
- They increase the risk of regulatory penalties as governments worldwide, including through the EU AI Act, tighten controls on biased AI systems.
- They contribute to commercial risks, poorly informed marketing strategies, product failures, and damage to brand reputation.
- Moreover, practitioners should strive to ensure their work doesn't knowingly reinforce systemic bias and discrimination.

SECTION 5:

Existing AI Ethics and Fairness Frameworks: From Macro to Operational

There are a number of existing frameworks and guidelines that provide vital moral and practical concepts that can support our thinking in this area. The main conclusion from a review of these documents is that none of them address the very specific requirements of the research sector. It is for this reason that a practical framework for Inclusive AI in research is needed.

1. Luxembourg Declaration on Artificial Intelligence and Human Rights[8]

“AI must be developed and deployed in full respect of fundamental rights and ethical principles, to ensure it serves human dignity and freedom.”

Scope: A European human rights framework emphasising the protection of fundamental rights, non-discrimination, transparency and democratic governance in AI development.

Relevance: Ethical baseline that highlights inclusion and fairness as human rights imperatives, critical to the research sector’s commitment to representing diverse voices.

Limitations: Primarily aspirational with limited operational guidance for implementation in industry contexts.

[8] https://humanists.uk/2025/07/10/humanists-pass-global-declaration-on-artificial-intelligence-and-human-values/

2. OECD AI Principles[9], EU AI Act[10] and America's AI Action Plan[11]

"AI systems should be designed to respect the rule of law, human rights, democratic values, and diversity, ensuring a trustworthy, fair, and inclusive society."
OECD

Scope:

- The **OECD AI Principles** provide broad, internationally recognised values emphasising trustworthy, human-centred AI, endorsed by over 40 countries.
- The **EU AI Act** is a comprehensive regulatory proposal setting risk-based requirements for AI systems, focusing on transparency, accountability, and protecting fundamental rights within the EU.
- **America's AI Action Plan** focuses on fostering innovation and national competitiveness, with some regulatory and governance ambitions.

Relevance:

These frameworks offer a macro-level foundation for ethical AI, emphasising risk management, bias mitigation, and transparency which are core concerns for market research AI applications.

Limitations:

- OECD principles are non-binding and broad without sector-specific guidance.
- The EU AI Act's complex regulatory scope may challenge practitioners, and it offers limited focus on systemic inclusion or equity beyond legal safeguards. This legislation is also being reviewed by the EU and is likely to change in 2026.
- America's AI Action Plan has faced criticism for insufficient emphasis on equity and civil rights protections, highlighting the need for sectors like research to lead in embedding inclusion proactively.

[9] https://www.oecd.org/en/topics/sub-issues/ai-principles.html
[10] https://www.oecd.org/en/topics/sub-issues/ai-principles.html
[11] https://www.oecd.org/en/topics/sub-issues/ai-principles.html

3. Corporate AI Ethics Frameworks

(e.g., Google AI Principles[12], Microsoft Responsible AI[13], IBM Fairness 360 Toolkit[14])

"Fairness is a measurable and actionable attribute of AI models that can be tested and improved."
IBM Fairness 360

- **Scope:** These offer practical ethical guidelines, toolkits, and technical resources designed to help developers and users identify and mitigate bias, enhance transparency, and ensure accountability.
- **Relevance:** Provide operational tools practitioners can adapt to conduct fairness audits, bias detection, and transparency assessments for AI-driven research methodologies.
- **Limitations:** Tech focussed and proprietary orientation means these may not fully address broader systemic inclusion issues or the specific ethical complexities of research contexts.

[12] https://ai.google/principles/
[13] https://www.microsoft.com/en-us/ai/responsible-ai
[14] https://research.ibm.com/blog/ai-fairness-360

4. Other Industry Specific Frameworks:

Financial Industry AI Ethics Frameworks (e.g., FINRA, European Banking Authority)

"Financial services must ensure AI systems are transparent, explainable, and operate free of bias to protect consumers and market integrity." FINRA

- **Scope:** Focused on governance, risk management, and explainability for AI in automated decision-making processes impacting consumers' financial wellbeing.
- **Relevance:** Demonstrate rigorous frameworks for trust, fairness, and accountability that the research sector can emulate to safeguard data integrity and ensure representative insights.
- **Limitations:** Tailored to finance, some principles may be less directly applicable to the nuanced ethical and inclusion challenges in research.

4. Healthcare AI Ethics Frameworks

(e.g., WHO Ethics and Governance of AI for Health)

"Ethical AI in health must prioritize equity, inclusiveness, and respect for individual rights to improve health outcomes for all." WHO

- **Scope:** Addresses the use of AI in healthcare with a strong focus on vulnerable populations, informed consent, privacy, and bias mitigation.
- **Relevance:** Shares parallels with research in protecting sensitive data and designing inclusive studies that consider intersectionality and diversity.
- **Limitations:** Healthcare-specific regulations and risks mean some guidance may require adaptation to fit research contexts.

While these frameworks provide critical foundations and useful tools, none fully address the unique ethical, technical, and inclusion challenges facing AI-powered research. Our sector requires a tailored framework that:

- Embeds **systemic inclusion by design** to ensure research insights genuinely reflect the full spectrum of society.
- Balances innovation and ethical responsibility, enabling AI to be a force for equitable insight generation rather than perpetuating bias.
- Equips practitioners with practical, actionable guidance specific to their workflows and challenges.

This is why developing a bespoke **Inclusive AI Framework** is essential not only to fill existing gaps, but to position the research sector as a proactive leader in ethical AI adoption.

SECTION 6:

Systemic Inclusion by Design: A Framework

The Systemic Inclusion by Design Framework has been created for research practitioners and organisations to utilise in the development of their AI systems.

This framework operates at the **AI system level**, embedding inclusion across data, models, and tools ensuring equitable outcomes from design to deployment.

An AI system is the whole ecosystem that makes artificial intelligence work not just the algorithm. It includes:

- Data: the information collected and prepared for learning
- Models: the algorithms trained on that data to find patterns or make predictions
- Tools: the applications or platforms that let people use those models
- Governance and feedback: the processes that monitor, update, and improve the system over time

To ensure it is as practical as possible, it is structured around the existing stages of AI system development. It does five things:

1. Articulates the various stages of AI system development. In this case 1-5.
2. It describes the Inclusion Framework Stage
3. It highlights the risks of not considering inclusion at this stage
4. It connects these risks back to fundamental ethical principles
5. It suggests tools and processes to use that address the inclusion risks

N.B.

There are a long list of potential tools and processes that could support this framework and inclusion by design for AI systems. Within this paper we describe three that are priorities (highlighted in the red boxes below) in the short-term, with the view to provide additional tool design in the future.

SYSTEMIC INCLUSION BY DESIGN: A FRAMEWORK

STAGE	INCLUSIVE AI FRAMEWORK STAGE	INCLUSION RISKS	EDI PRINCIPLES AND VALUES	INCLUSION TOOLS / PROCESSES
1	Systemic Inclusion by Design	Narrow framing of the research objective Exclusion of diverse perspectives	Representation Equity of voice Power awareness	1. Systemic Inclusion Canvas: values alignment, stakeholder mapping, context setting 2. Inclusive Design Sprint Toolkit: participatory design workshops for upfront equity framing 3. *Future Tool*: Impact Scoring Matrix for inclusion risk forecasting at project kick-off
2	Inclusive Data Practices	Sampling bias Exclusion of protected groups Label bias and language skew	Fair access Cultural sensitivity Transparency	1. Inclusive Sampling Framework: oversampling strategies for underrepresented communities 2. Bias-Aware Labelling Guide: standards for fair annotation and contextual metadata 3. *Future Tool:* Provenance Tracker for annotator demographics, language variance, and consent logs
3	Equitable Model Development	Proxy variables encoding bias Training data homogeny Algorithmic opacity	Algorithmic fairness Intersectionality Accountability	1. Bias Audit Protocols (e.g., IBM AI Fairness 360): scoring models for fairness across groups 2. Demographic Performance Reporting: surfacing differential outcomes across identities 3. *Future Tool*: Equity-weighted loss functions tailored to survey-based prediction tasks
4	Transparent Implementation & Communication	Lack of explainability Unclear limitations Misalignment with diverse users	Informed consent Accessible design Trustworthiness	1. Model Scoring Cards: plain-language disclosures of model assumptions and uses 2. Inclusive UX Standards: ensure usability across literacy, language and ability levels 3. *Future Tool:* Dynamic Transparency Layer – contextual explanations adapted to user needs
5	Continuous Accountability & Improvement	Bias drift over time Ignored community feedback No reparation for harms	Co-creation Ongoing inclusion Responsive governance	1. Inclusive Feedback Loops: structured listening from impacted communities post-deployment 2. Post-launch Bias Monitors: metrics that detect drift or exclusion over time 3. Future Tool: Community Audit Network: third-party, lived-experience-led auditing model

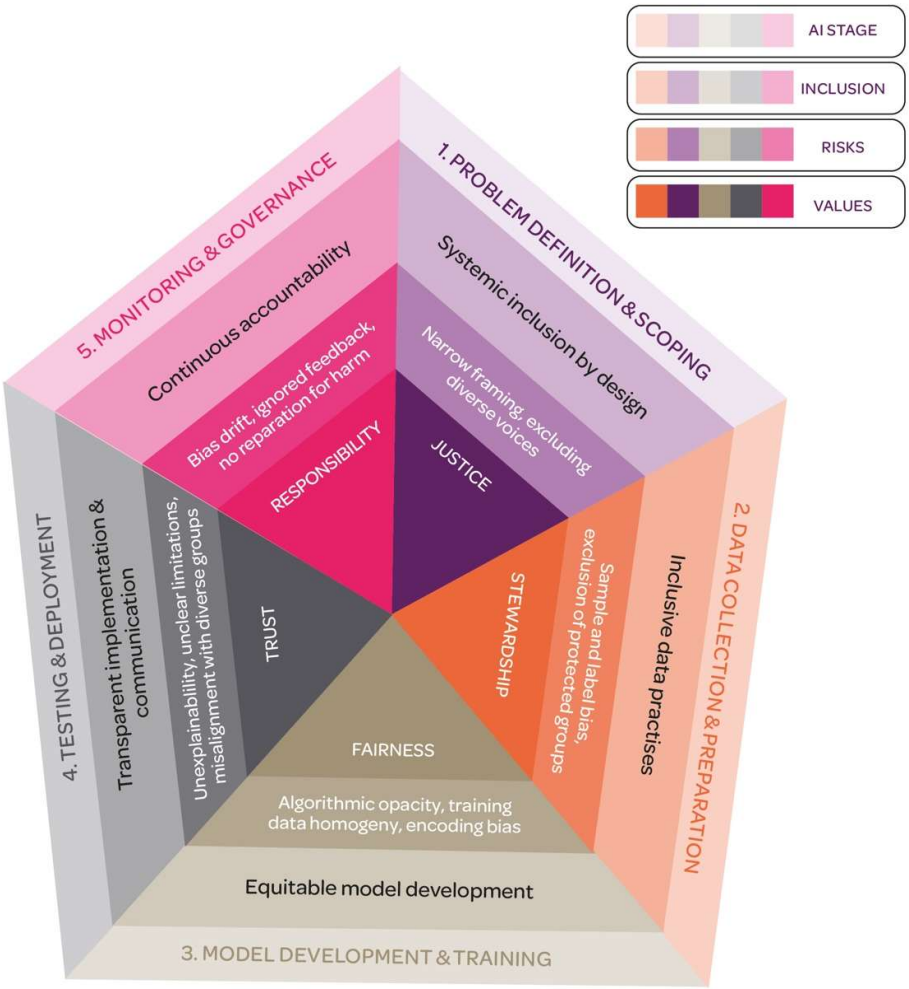
We have also illustrated the Systemic Inclusion by Design Framework as a prism, a reminder that building responsible AI requires looking at your system from multiple angles

Each face represents a stage of development, its inclusion focus, the risks it carries, and the human values it protects.

As practitioners move around the prism, their perspective shifts, revealing new facets of bias and impact. And because the prism loops back on itself, it shows that inclusion isn't a one-off task but a continuous cycle of reflection and refinement.

The prism helps teams see their AI through **the lens of inclusion**, making risks clearer and human values harder to ignore.

Prism of Systemic Inclusion by Design



SECTION 7:

Inclusion Tools in more detail

There are a number of potential Inclusion Tools applicable to each stage of development, however this paper focuses on three specific tools that can deliver the most impact for least effort and arguably should be addressed first.

Stage 1: Systemic Inclusion Canvas

Designed to act as a first-stage inclusion scoping prompt for all new and existing tools, helping teams surface equity risks, map affected stakeholders, and align design with organisational values.

Stage 3: Bias Audit Protocols

Structured guidelines for identifying, measuring, and mitigating algorithmic bias across the data and model lifecycle, ensuring fairness checks are embedded in technical QA processes.

Stage 4: Model Scoring Cards

A practical evaluation tool to assess the inclusivity of datasets, model outcomes, and interfaces offering transparent, repeatable metrics that can sit alongside traditional audit frameworks.

SECTION 8:

Stage 1 Tool
Systemic Inclusion Canvas

The **Systemic Inclusion Canvas** is fundamental to the entire process of developing AI systems. It's a working document designed to provide integral opportunities for discussion at the first stages of AI design and is based on Olivia Gambelin's ethical AI methodology [15]which applies a value-led evaluation loop to the stages of development.

Suggested guidance for use:

- Teams should approach this document with honesty, collaboration and look to problem solve rather than justify or defend.
- The structure can be adapted to fit the needs of the business but provides momentum to avoid overwhelm and stagnation.
- A facilitator should be assigned to chair the regular meetings – someone with interest or familiarity with inclusive principle or ethical AI
- A variety of participants should contribute, such as technical, research, senior leadership and inclusion.
- Use the document as a guide, surface key insights, highlight high-risk areas, gaps, or overlooked stakeholders.
- Using a RACI framework can help to clarify tasks and decisions:
    - **Responsible:** The individual(s) who perform the work to complete the task.
    - **Accountable:** The one individual who has the final authority and signs off on the completed task.
    - **Consulted:** Individuals whose expert input is required before a decision is made or the task is finished (two-way communication).
    - **Informed:** Individuals who must be notified after the task is completed or a decision is made (one-way communication).

[15] https://www.thevaluescanvas.com/

- Not all tasks can be completed immediately, set priorities for the interventions based on impact.
- Revisit the canvas on a regular basis, and treat it as a living document that evolves as systems and context changes.

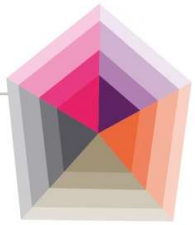
Detailed description of Systemic Inclusion Canvas:

Canvas Section	Aligned Ethical AI Stage	Key Prompts (for AI tools)	Market Research: Real Life example
1. Systemic Inclusion Proposition	Values Discovery	What inclusive values guide the tool's purpose? What systemic biases are we aiming to mitigate in the research ecosystem? Are we solving a real equity problem or just scaling convenience?	An AI survey writing assistant is built to reduce bias in question design. Its core principle is to reflect intersectional participant experiences, not just speed up scripting.
2. Stakeholder Segments	Stakeholder Mapping	Who uses this tool? (researchers, clients, participants?) Who is impacted downstream by its outputs? Are marginalised identities represented in testing, data and development?	A segmentation model builder does not account for non-binary gender leaving out non-conforming audiences in automated clustering recommendations.
3. Problem Definition	Values Discovery + Context Framing	What aspect of research is the AI improving? Is the tool trying to automate something that requires human cultural nuance? Does the framing reinforce outdated assumptions or power imbalances?	An AI trend analyser assumes Western online sources are representative of global sentiment. This skews insights and misses culturally relevant context.
4. Context & Environmental Framing	Context Framing	What social, political or global context does the tool operate in? Are training datasets pulled from biased or limited historical norms? Is the tool inclusive across geographies and communities?	A sentiment model trained on U.S. datasets misinterprets expressions of Black British English vernacular as negative tone.
5. Access & Engagement Channels	Stakeholder Mapping + Context Framing	Who has access to using or customising the tool? Does it require technical knowledge or language proficiency to be effective? Are small/under-resourced teams excluded from shaping or benefiting from it?	An AI platform's dashboard and outputs are only available in English, limiting accessibility for non-native-speaking practitioners or local language practitioners.

Canvas Section	Aligned Ethical AI Stage	Key Prompts (for AI tools)	Market Research: Real Life example
6. Ethical Safeguards & Design Solutions	Risk Anticipation	What inclusion guardrails are built in (e.g. demographic weighting, fairness flags)? Is explainability possible? Are there 'stop' points when the tool might produce harmful or inaccurate outputs?	An image recognition tool for ad testing includes a fairness audit step to detect low accuracy when analysing dark-skinned faces in visual creative testing.
7. Measurement of Inclusion & Fairness	Risk Anticipation	What metrics track fairness in output (e.g., demographic parity, bias reduction)? Are error rates monitored across identity groups? Are users aware of fairness benchmarks?	A predictive purchase intent model includes dashboards showing confidence levels by race/ethnicity and income segments to flag overfitting in skewed training data.
8. Integrity Advantage & Decision Review	Decision Review	Were inclusive decisions undertaken through development? Were users/clients given clear explanations of limitations? Was conduct inclusive user testing undertaken before deployment?	The beta version of a participant-moderation AI tool was paused after non-white faces were disproportionately flagged. User feedback triggered retraining and public disclosure.
9. Sustainability & Future Impact	Risk Anticipation + Context Framing	What's the environmental cost of this tool (compute, scale)? Could the tool embed long-term harm (e.g. stereotyping, exclusion)? Will it require regular audits to stay fair and relevant?	An LLM-based open-end coding tool increases compute demand dramatically. Designers include lightweight local fallback models for small-scale practitioners to reduce emissions.

***Suggested template document** (see full design in the Appendix)

Systemic Inclusion Canvas



1. Systemic Inclusion Proposition

- What inclusive values guide the tool's purpose?
- What systemic biases are we aiming to mitigate in the research ecosystem?
- Are we solving a real equity problem or just scaling convenience?

1. Equity and Representation: Making sure all relevant groups are included
2. Bias awareness: Explicitly avoiding stereotypes
3. Transparency: Users understand how outputs are generated
4. Actionable and responsible insights: that don't perpetuate harm

2. Stakeholder Segments

- Who uses this tool? (researchers, clients, participants?)
- Who is impacted downstream by its outputs?
- Are marginalised identities represented in testing, data and development?

1. Equity ecosystem

2. Problem Definition

- What aspect of research is the AI improving?
- Is the tool trying to automate something that requires human cultural nuance?
- Does the framing reinforce outdated assumptions or power imbalances?

1. Equity ecosystem

1

SECTION 9:

Stage 3 Tool
Bias Audit Protocols

This tool is an audit protocol to mitigate the algorithmic bias found in new and existing AI systems.

Algorithmic biases can be created in a huge number of ways, and they not only create inaccurate outputs but also risk perpetuating discrimination and biases for individuals.

- **Data Biases** e.g. Representation, Sampling, Historical, Measurement, Labelling/Annotation. For example: Health tracking apps trained mostly on young, healthy volunteers OR predictive policing models trained on historically over-policed neighbourhoods OR human annotators injecting their own cultural biases onto images.

- **Proxy Bias** e.g. Feature selection or proxy variables For example: Using Postcode/ Zip code as a proxy for credit-scoring.

- **Algorithmic / Modelling Bias** e.g. Optimisation objective bias, Model Design bias, Default Parameter. For example: Optimising for engagement, therefore amplifying extreme or divisive content OR Linear models can't capture nonlinear differences OR choosing default thresholds that might disadvantage certain groups.

- **Human AI Bias** e.g. Feedback loop, interface and Framing bias. For example: Predictive policing, more police are sent, more recorded incidence, model 'learns' it's a high crime area

- **Deployment or Context Bias** e.g. Transfer or operational bias. For example: Model trained in one place, then transferred to another which has different performance and demographics

- **Social and Structural Bias:** e.g. Structural inequalities, Cultural Bias. For example: Assuming one cultural perspective as universal such as Western job titles, norms etc. OR credit models punishing applicants from historically redlined neighbourhoods.

Given the complexity there is not a 'one size fits all' to address these bias issues.

There are a number of existing technologies and audit practices that are available for teams to use, however they are not always tailored to the needs of research practitioners.

As such, this paper provides direction as to which existing techniques are potentially relevant and where supplement techniques could be considered.

From these tools, developers can 'Score' the tool on how biased it is and put in place actions to address the issues.

The benefit of aligning our approach is to create a consistent approach that is embedded in the workflow.

Stage	What to Audit	Relevant Techniques/Tools	Why It Matters in Market Research
1. Define Audit Scope & Risk Areas	Clarify the tool's use (e.g., segmentation, targeting, predictive modelling) and intended population.	1. ***Microsoft's Responsible AI Standard*:** [16]Emphasises stakeholder inclusion and impact assessment. ***2 Systemic Inclusion Canvas (Appendix)***	Ensuring your team question who is benefiting from the tool and who is being excluded, rather than only focussing on commercial outcomes
2. Identify Protected & Proxy Attributes	Check for features that might encode unfair bias directly (e.g., gender) or indirectly (e.g., postcode = income proxy).	1. **IBM's AIF360**[17]: Proxy detection & correlation tests. 2. **Google's Model Card** [18]guidance: Identify sensitive features.	For example, in research we might use behavioural data such as App usage, as a proxy for demographic or psychological traits.
3. Dataset Bias Check	Assess whether training or input data reflects the full population the research aims to understand.	**1. Google's Inclusive ML Practices**[19]: Evaluate sampling, oversample underrepresented groups. **2. Data Nutrition Labels** [20](MIT/Google initiative)	Unrepresentative datasets distort insights. Dataset checks protect against this for example, skewed brand trackers, faulty personas, or biased behavioural predictions.
4. Fairness Testing Across Demographics	Test whether model performance, outputs, or recommendations differ across identity groups.	**1. IBM's AIF360:** Test for disparate impact, equal opportunity. **2. Microsoft's Fairlearn[21]:** Dashboard to visualise model performance by group.	If the model predicts poorly for one group, insights may be misleading or even discriminatory.

[16] https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Responsible-AI-Standard-General-Requirements.pdf?culture=en-us&country=us
[17] https://ai-fairness-360.org/
[18] https://modelcards.withgoogle.com/model-cards
[19] https://developers.google.com/machine-learning/guides/rules-of-ml
[20] https://datanutrition.org/labels
[21] https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/

Stage	What to Audit	Relevant Techniques/Tools	Why It Matters in Market Research
5. Explainability & Interpretability	Ensure model logic is understandable to stakeholders.	1. **Google's What-If Tool** [22](in TensorFlow): Interactive fairness testing 2. **LIME and SHAP**[23]: Explain model decisions.	In research, clients and participants must trust AI outputs. Lack of explainability undermines credibility.
6. Transparency & Documentation	Clearly communicate limitations, assumptions, and audit findings.	**1. Google's Model Cards.** **2. Microsoft's Datasheets for Datasets.**[24] 3. **Plain-language executive summaries** (industry-specific process)	Transparency is essential for ethical claims and client accountability. Datasheets provide a repeatable audit step: record assumptions, limitations, and risk areas upfront.
7. Remediation & Action Planning	Decide how to address bias discovered in the audit.	1. **Audit Response Plans** (new industry tool): Pre-set thresholds for retraining, flagging models, or pausing rollouts.	Either an industry wide or organisation specific plan, post inclusion audit on your subsequent actions e.g. re-train model with balanced dataset or adjust scoring thresholds OR pause deployment, review with ethics/legal team
8. Ongoing Monitoring	Continue to audit for bias drift post-deployment.	1. **Bias Drift Dashboards** (new industry tool). **2. Scheduled Bias Checkpoints.**	Provide a **real-time or periodic visual overview** of model performance and fairness across key demographic, behavioural, and psychographic segments, so teams can detect bias or drift over time.

[22] https://cloud.google.com/blog/products/ai-machine-learning/introducing-the-what-if-tool-for-cloud-ai-platform-models
[23] https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models
[24] https://www.microsoft.com/en-us/research/project/datasheets-for-datasets/

SECTION 9:

Stage 4 Tool
Model Scoring Cards

At this point there is no standardised way to evaluate how inclusive an AI tool or system is within the research sector. And although in a lot of cases the models and tools differ from one another, it is arguably important to have a common language, and high-level dimensions that are relevant regardless of the approach.

The Inclusion Scoring Card is a starting point in creating some standardisation across systems, that allow clients, stakeholders and non-technical teams to understand how a tool performs on an inclusive evaluation.

The idea is that the core dimensions, scoring and responsible teams will probably stay the same, but there may be bespoke metrics that fit within each dimension depending on the type of AI system.

Why this is useful:

- Standardising transparency
- Creating accountability
- Supporting with client confidence
- Supporting accessibility outside of deep AI knowledge

How it works:

- A nominated owner should coordinate the evaluation across the various dimensions; in theory this could be anyone but if they have some understanding of Responsible AI that is an advantage.
- The card should be completed for each tool, and again if a tool is new or significantly updated.
- There is a shared responsibility to support the evaluation and action points without penalising honesty.
- This card could be published alongside the project deliverables or within the tool documentation or RFQ submissions.

AI Model - Inclusion Scoring Card

	DIMENSION	DESCRIPTION	METRIC	SCORE	TEAM	DATE
1	REPRESENTATIVENESS AND INTERSECTIONALITY	Training data representative of target population and includes unrepresented or overlapping identity groups	Coverage % by demographic groups, intersectional groups, comparison to population benchmarks, notes on gaps		▪ Data Science ▪ Research Analytics	FEB 2025
2	ACCESSIBLE AND CULTURALLY SENSITIVE	Ensure model outputs are culturally appropriate, understandable and accessible through the interface	Language coverage, misinterpretation rates, review of cultural bias, readability, navigation, accessibility scores (WCAG), clarity rating 1–5		▪ UX ▪ Design ▪ Researchers or EDI Team ▪ Accessibility Team	JUNE 2024
3	FAIRNESS OF OUTCOMES	Measure whether predictions, classifications or recommendations are equitable across groups	Disparate impact ratio, equal opportunity, false positive/negative rates by group. Use IBM AIF360 / Fairlearn metrics or similar		▪ AI / ML Development ▪ Data Science	JUNE 2024
4	TRANSPARENCY AND EXPLAINABILITY	Can stakeholders understand how the model produces outputs?	Provide model cards, datasheets and give a clarity rating 1-5		▪ AI Team ▪ Responsible AI	FEB 2025
5	CONSENT AND PRIVACY	Ensure data use aligns with participant consent and legal/ ethical standards	GDPR compliance, and MRS Code of Conduct compliant. Anonymisation and provide audit notes		▪ Data Governance ▪ Ethics ▪ Responsible AI	AUG 2025
6	ETHICAL AND SOCIETAL IMPACT	Assess potential harms, unintended consequences and systemic effects	Risk Rating (low/medium/high) on potential societal / community impacts		▪ Project management ▪ EDI/ Responsible AI Team	AUG 2025
7	ACTION AND REMEDIATION	Documents next staps to improve inclusivity, fairness, accessibility or reduce risk	Specific action points, assigned owners, timeline for remediation		▪ Cross functional teams ▪ Project management	ONGOING

Conclusion

AI systems are being developed at break-neck speed and across every corner of the research sector; this paper aims to give space for pause and assessment of those systems. Asking the fundamental question, are these systems fair, and who are they fair for?

The risks of not working in an 'Inclusion by Design' way include biased outputs, client distrust, reputational harm, operational inefficiency and missed opportunities.

Inclusion by Design is all about increasing the accuracy and accountability of what we are building and creating a long-term strategic advantage.

This paper highlights the need for organisations and practitioners to understand, familiarise and implement an Inclusion Framework to guide the development of your AI systems.

- Provide research practitioners and organisations within the research sector, a tailored framework to drive inclusive AI development.
- Support activation with three important tools that can be utilised immediately.
- Provide a suggested implementation guide to utilising the framework and tools.

The Three suggested Tools

Within the Inclusive AI Framework, we explored three specific tools that will be most useful immediately.

1. **Systemic Inclusion Canvas**: This working document should be used at the very start of AI ideation, and consulted throughout the development process, it aims to be the Inclusive Voice in the Room to help guide inclusive-first design.
2. **Bias Audit Protocols:** Specifically, to test the biases prevalent in systems throughout the development process, practical tools to understand where the issues are and what to do about them
3. **Inclusion Scoring Card**: High level and potentially externally facing standardised scoring of how inclusive your tool is. This can be used to support transparency and accountability internally and externally.

AI Ethics Committee

It's likely that a large proportion of the team developing your AI tools and involved in decision making are technical engineers, data scientists or senior leadership. What is often missing is an Ethical, Responsible or EDI voice at the table.

One of the most effective ways to embed an inclusion-by-design approach is to establish an AI Ethics Committee. This should be a diverse, cross-functional group, ideally with a **Responsible AI champion** who can guide practice, not just policy.

The role of this committee is simple but crucial: to own the Inclusion Framework, drive action, and act as the organisation's sounding board for ensuring AI tools are fair, representative, and safe before they reach clients or participants.

As well as individuals within an organisation, it can also include external voices, such as ethics advisors, or members of the community who will be impacted by your tools to collaborate in the development.

"Good technology does not take advantage of our human nature, it's tech that helps us embrace the nature of being human." Olivia Gambelin[25]

Reach out to the AIA Council for more information on this paper

[25] www.oliviagambelin.com/

Appendix

Systemic Inclusion Canvas Template

Systemic Inclusion Canvas

1. Systemic Inclusion Proposition

- What inclusive values guide the tool's purpose?
- What systemic biases are we aiming to mitigate in the research ecosystem?
- Are we solving a real equity problem or just scaling convenience?

1. Equity and Representation: Making sure all relevant groups are included
2. Bias awareness: Explicitly avoiding stereotypes
3. Transparency: Users understand how outputs are generated
4. Actionable and responsible insights: that don't perpetuate harm

2. Stakeholder Segments

- Who uses this tool? (researchers, clients, participants?)
- Who is impacted downstream by its outputs?
- Are marginalised identities represented in testing, data and development?

1. Equity ccccccccc

2. Problem Definition

- What aspect of research is the AI improving?
- Is the tool trying to automate something that requires human cultural nuance?
- Does the framing reinforce outdated assumptions or power imbalances?

1. Equity acccccccc

1

Systemic Inclusion Canvas

4. Context & Environmental Framing

- What social, political or global context does the tool operate in?
- Are training datasets pulled from biased or limited historical norms?
- Is the tool inclusive across geographies and communities?

1. xxxxxxxxxxxxxx

5. Access & Engagement Channels

- Who has access to using or customising the tool?
- Does it require technical knowledge or language proficiency to be effective?
- Are small/under-resourced teams excluded from shaping or benefiting from it?

1. Equity ccccccccc

6. Ethical Safeguards & Design Solutions

- What inclusion guardrails are built in (e.g. demographic weighting, fairness flags)?
- Is explainability possible?
- Are there 'stop' points when the tool might produce harmful or inaccurate outputs?

1. Equity acccccccc

2

Systemic Inclusion Canvas

7. Measurement of Inclusion & Fairness

- What metrics track fairness in output (e.g., demographic parity, bias reduction)?
- Are error rates monitored across identity groups?
- Are users aware of fairness benchmarks?

1. xxxxxxxxxxxxxx

8. Integrity Advantage & Decision Review

- Were inclusive decisions undertaken through development?
- Were users/clients given clear explanations of limitations?
- Was conduct inclusive user testing undertaken before deployment?

1. Equity ccccccccc

9. Sustainability & Future Impact

- What's the environmental cost of this tool (compute, scale)?
- Could the tool embed long-term harm (e.g. stereotyping, exclusion)?
- Will it require regular audits to stay fair and relevant?

1. Equity acccccccc