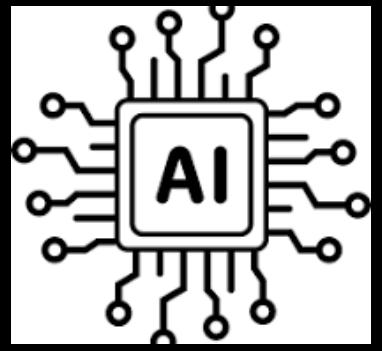
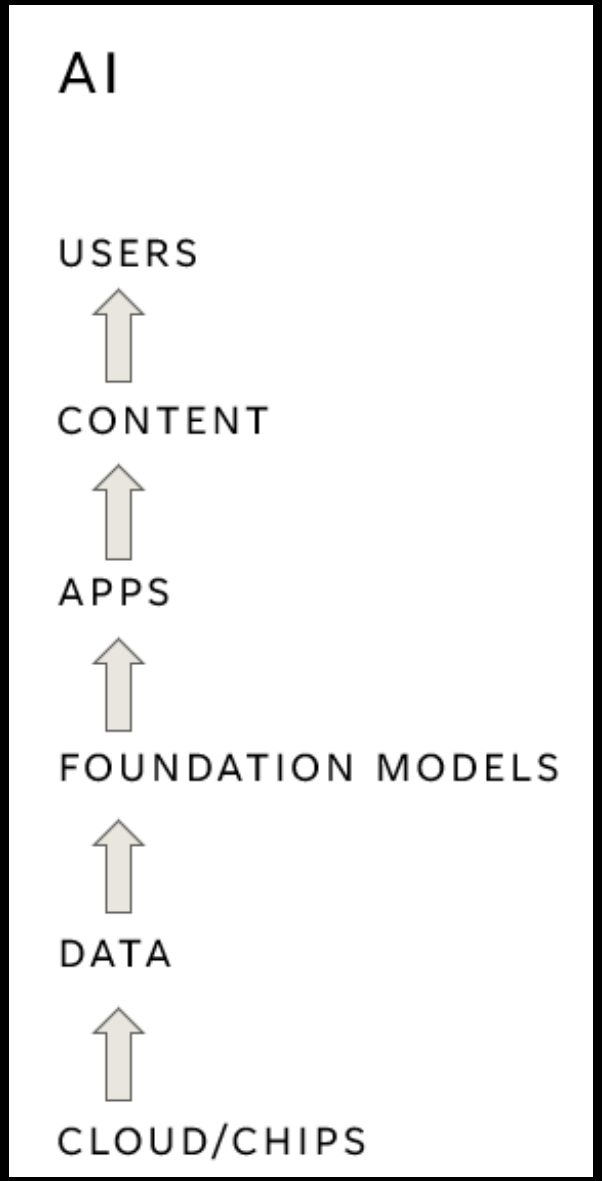
EMB Strategic Communications in the Context of the Digital Media: harnessing the opportunities of AI in electoral processes



Guy Berger

AI tech stack: for Generative AI & Classificatory AI



For African elections integrity, what matters are the features in:

- Foundational data
- Foundational models
- The Apps used
- The Content layer, which rests on addtional data inputs to the App (eg. commands; your documents), and the and/or results thereof

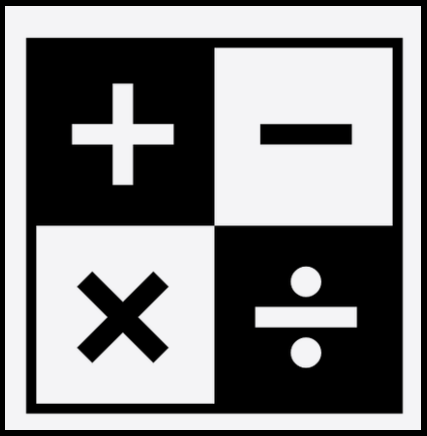
The AI opportunities/risks calculus

Every (AI) opportunity also entails risks

Beware of false “balancing”:

Some risks can wipe out the benefits – it’s not an average

Remember some risks create opportunities for mitigation



The opportunities/risks calculus

Point of view: “the election” & its actors

AI use for EMBs as custodians

AI use among political contenders

AI use by media, observers, civil society, security, researchers, etc.

AI use is not in silos: use in one block impacts on the others – and therefore the whole

1. Managing the AI security paradox

Risk: digitalisation and AI opens new cybersecurity vulnerabilities for all actors

Opportunity: AI can help enhance election security and detect AI-fuelled attacks

Mitigation: Avoid a single-point of failure –have multi-layered security systems;

Use AI to help with threat assessment and monitoring

2. The 2-edged sword: AI personalization

Opportunity for personalised info to target specific voters - both in Voter Education, and by political campaigners…

Risk: Public backlash.

Potential exploitation of data for nudging (e.g. vote suppression)

Mitigation: EMB practice that prioritizes voter agency,

Rules on transparency on targeting, limits on micro-targeting capabilities

3. Optimising EMB efficiency gains

Opportunity: analyse statistics, summarise documents, translate – inputs for EMB comms

- **Mitigation**: Use of tools requires transparency
- Require vendors to show algorithmic audits, diverse data, and copyright respect

Risk:

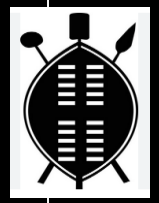
- Potential systemic biases
- Confidentiality concerns
- Snake-oil vendors

4. AI for EMB comms – pro's and con's

Opportunity: Generative AI chatbot interfaces to respond to specific voter queries

Risk: AI powered systems can produce false answers

- **Mitigation**: Damp down Generative AI for chatbots, limit the possible prompts, stick to fixed FAQ style answers; channel other queries to human intelligence



AI-based comms for polls

- *Generative AI relies on predictive word association patterns in its database*, meaning that:
- AI-generated content can <u>often be incorrect</u>, even irrespective of whether the prompters intended this or not

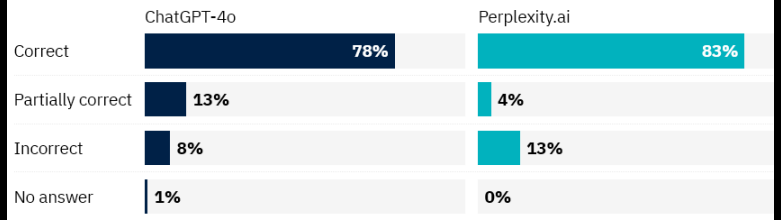
Seeking Reliable Election Information? Don't Trust AI

Experts testing five leading AI models found the answers were often inaccurate, misleading, and even downright harmful

By Julia Angwin, Alondra Nelson, Rina Palta
Feb 27, 2024

Percentage of election-related questions for which chatbots provided a correct answer

ChatGPT-4o and Perplexity.ai provided correct answers to the majority of questions, but also provided partially correct and incorrect answers.

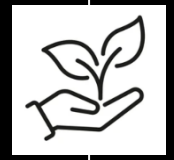


A clear majority of responses from ChatGPT-4o and Perplexity.ai were overall accurate, but we saw some notable errors, such as incorrect listings of candidates. Some of these inaccuracies seemed to stem from the chatbots' difficulties in handling tasks that require reasoning about and contextual understanding of the questions. Nevertheless, the incorrect answers were stated with the same certainty as responses that were correct. The chatbots frequently

5. Tackling AI-generated disinformation

Risk to all actors: AI can generate sophisticated disinformation (*tempting some actors to use it*)

Opportunity: AI can partly help to detect and flag some fakes, and discern networks of bots and trolls.



- **Mitigations**: Code of conduct for political parties.
- EMBs & fact-checkers monitor for AI-generated false content.
- Platforms are asked to fulfil their commitments…

Deep dive: AI threats to Info Integrity

Caveats about AIGC panic:

- *Misleading content does not have to be factually false, just out of context*
- *Faked content can be made without using AI as such…*

NEVERTHELESS:

- AI tools radically cheapen disinfo production and its targeting.
- Synthetic AI-generated content is new, it's not like existing images that can be reverse-searched

Proliferation ahead = “  ”?

AFP Fact Check

Share:

South Africa elections 2024 | Artificial intelligence

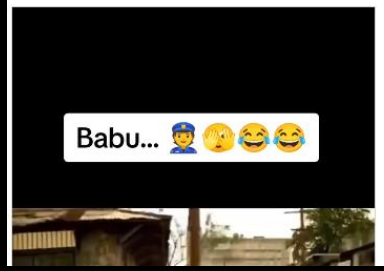
Doctored video does not show Trump backing new South African party

SOUTH AFRICA • 5 JUNE 2024

AI-generated video mocks outgoing Minister of Police

ELECTION | 29 MAY 2024

South Africa • Video • X/Twitter



SOUTH AFRICA • 8 MARCH 2024

In fake video, U.S president threatens sanctions on South Africa

ELECTION | 29 MAY 2024

South Africa • Video • X/Twitter • Facebook • TikTok

AI-generated potholes warn off South African voters

ELECTION | 29 MAY 2024

South Africa • Image • Facebook • X/Twitter • Instagram • Reddit

Home / 2024 / Setembro / 16 /
Chissano's alleged praise for Venancio Mondlane is fake

English

Chissano's alleged praise for Venancio Mondlane is fake

Paul Fauvet | 2024-09-16 | 2 min read

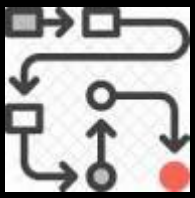
Home | Watch | Listen | Play | Publications

To further their appeal with young people, politicians like Mondlane, who is 50, or Botswana's new president Duma Boko, who is 54, have also tried to modernise their appearance.

Mondlane has grown his hair -- a style mocked by Frelimo supporters who publish fake pictures of him holding a shampoo bottle -- and Boko, who dons a clean-shaved haircut, has shared multiple photos of him in sports clothes.

"Optics matter in politics," said Mbanje. "Political figures looking young will always resonate with young people."

Distribution – the biggest problem?



- AI can supercharge *convincing* image, audio, video…
- And can easily customize text for different *target* audiences
- But to be effective, outputs then need to be distributed!
- *Enter the social media platforms* (and low quality news media)

- …where co-ordinated inauthentic networks work to exploit virality logics
- Platforms *could* do a lot to –
  - Stop *algorithmic* amplification
  - *Improve* the use of Classification AI (in relevant languages) to better moderate harmful content and sanction accordingly.

Tackling the risk of AI generated disinfo

An EMB will need to prioritise:

- Which manipulated media challenges to Info Integrity merit what mitigations?
- What mitigations are most feasible *and* most effective…
- (Keeping in mind also that mitigations go much wider than only direct responses to AIGC)

Defining harms (and at what priority?)

1. Fakes of candidates and their statements/policies
2. Fakes of endorsements
3. Fakes on voting arrangements and results
4. Plagiarism of party symbols and webpages.
5. False depiction of events
6. Impersonation of EMB officials
7. Fakes about EMB officials

Where to expect AI-disinfo?

Is it:

- on social media, social messaging, mainstream media?
- mainly visual, audio, video, text?
- equally in campaign materials and in political advertising?
- targeting mainly once-off content or trending content?

Be prepared:

- Assess risks and do a playbook for scenarios

Risk template: High, medium or low? A, B, C?

Type of risk identified for elections	Impact	Likelihood	Priority
To freedom of expression:			
• Silencing online public voices by intimidation			
• Conspiracies communicated on WhatsApp			
• Incitement to violence via social media			

Assess risk by looking at impact (severity) and likelihood (probability)

Matrix based on SA National Editors Forum and used in the SA election

Risk template for AI-created electoral disinfo:

Type of risk identified for elections	Impact	Likelihood	Priority
To access to information:			
• Disinformation on election processes			
• Manipulated media			
To electoral integrity:			
• Online attacks on electoral integrity / results			
• Hacking and impersonation of ECZ social media presence			
Other risks			

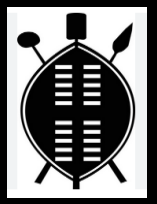
In the case of AI-fuelled electoral disinfo:

Type of risk identified for elections	Impact	Likelihood	Priority
To access to information:			
• Disinformation on election processes	High	Low	B
• Manipulated media	Medium	Low	B
To electoral integrity:			
• Online attacks on electoral integrity / results	High	High	A
• Hacking and impersonation of the EMB's own social media presence	Low	Low	C
Other risks			

Identify possible mitigations

Type of risk identified for elections	Who should prepare, and how?	Trans-parency of response?	Priority
To access to information:			
• Disinformation on election processes	Actors ii, iii, iv	Yes	B
• Manipulated media	ii, v, viii, etc	Yes	B
To electoral integrity:			
• Online attacks on electoral integrity / results	Etc	Yes	A
• Hacking and impersonation of the EMB's own social media presence	Etc	Yes	C
Other risks			

AIGC mitigation roles for actors:



- Tech companies to step up – *but don't count on them to do so!*
- Election contenders' ethics; and interest in policing each other!
- Civil society and researchers to help detect and monitor
- Law enforcement to investigate (*though tracking is tough*)
- Courts (to rule where needs be)

- Your own role – a code of conduct with contenders? (and influencers and platforms?)

Implementation:
Can violations be found & verified?

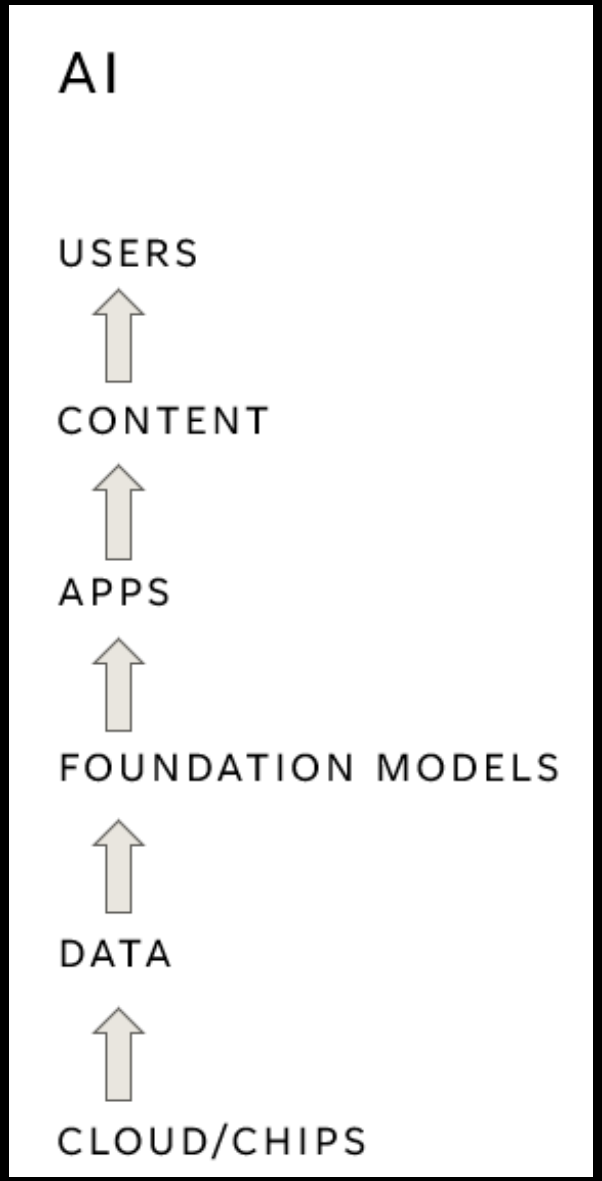
***Not all**! But create some friction!*

So an EMB will need to:

- Prioritize kinds of cases and places for monitoring and enforcement…
- Create – and publicise – identification mechanisms:
  - Contender-source & crowd-source egregious cases of the 7 instances
  - Set up rapid response capacity
- Proactively monitor & respond to what YOU know is false.
- Create partnerships with fact-checkers for other cases…

Ask platforms to help identify original posts and extent of co-ordinated inauthentic networks – i.e. to assist in attribution

Access to data is key



Garbage in,

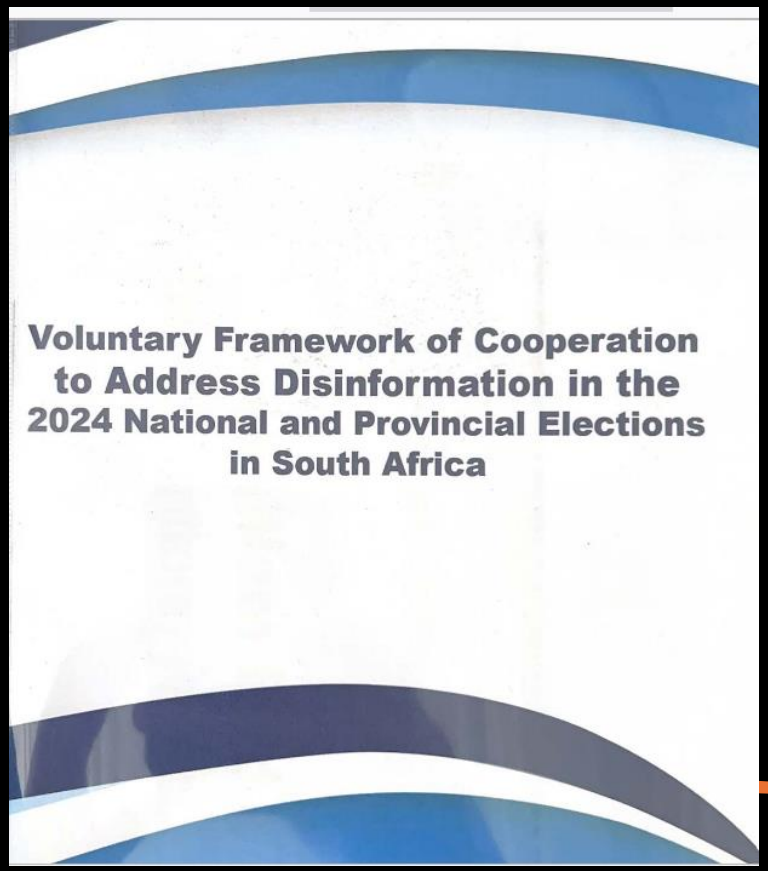
garbage out

Getting the right data! Whether from the census, social media monitoring, weather…

The sources are key!

Without Africa unlocking relevant data, AI systems are highly limited..

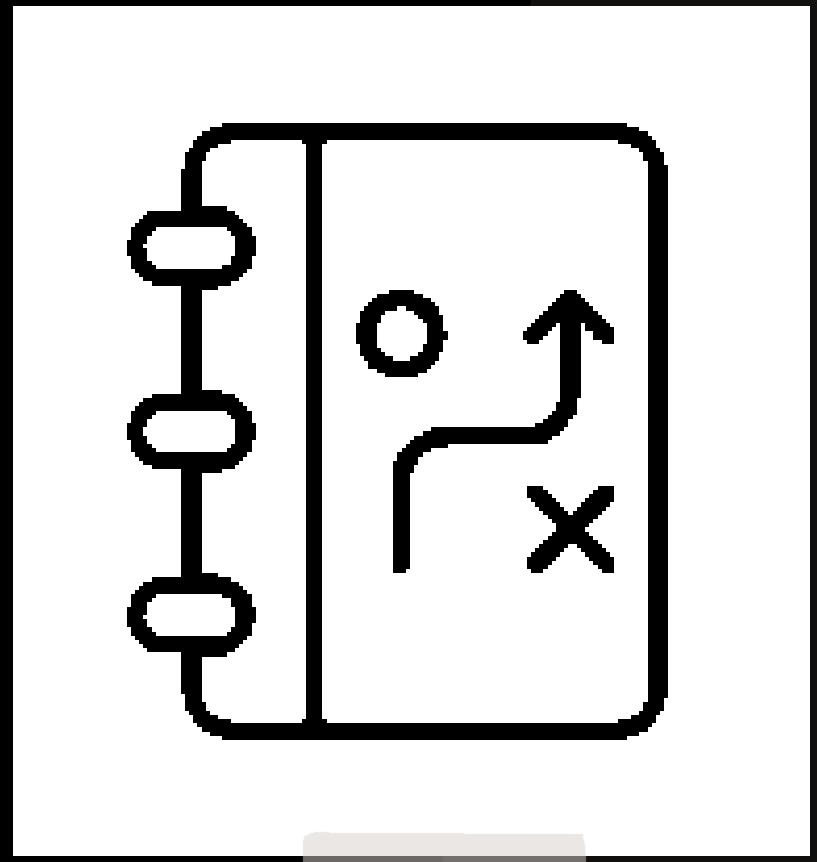
Partnerships and progress



Recap: Storm warning: AI-fuelled manipulation ahead

- *Prioritisation will be needed as to your use of, and response to, AI.*
- *Consider a human rights risk assessment*

Playbook 1



Examine likely opportunities and harms:

Checklist: Bias, inaccuracy, confidentiality

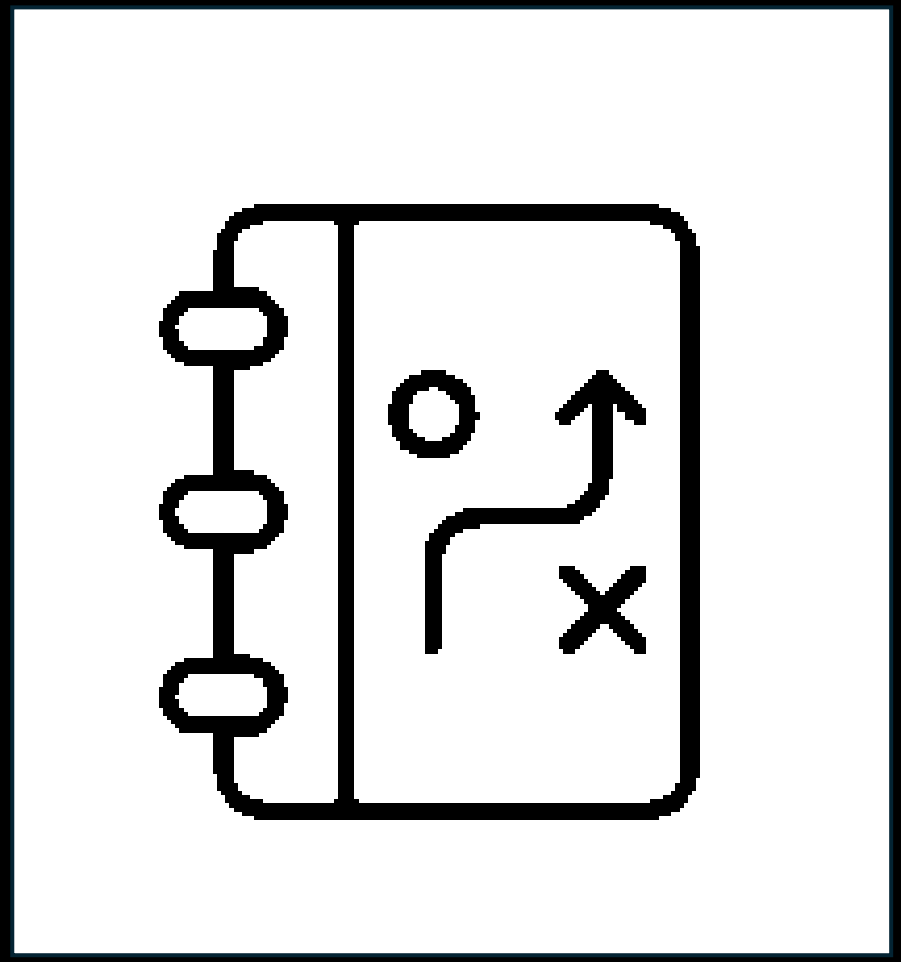
Checklist: what kinds of rules for use of manipulated media (bans, labels, other)

- For all the above, keep in mind what is feasible

Create systems for monitoring, reporting and responding

Give attention to data!

Playbook 2



<u>On the enabling side</u>, for Info Integrity:

- Make info available as a public good & communicate this…
- Enhance EMB transparency (incl data)
- Focus on data quality and representativity
- *Caution on data confidentiality*
- Do partnerships, talk to tech
- *Raise the AI literacy of staff, electoral contenders, police, courts and the public!*

All this on top of your day-jobs!

Thank you

I write on these issues on LinkedIn

Contact: G.Berger@ru.ac.za