The Pitfalls of AI

What are the implications of the development of AI for leadership in government, business, research, and society?

FULUFHELO NELWAMONDO, PHD

CEO: NATIONAL RESEARCH FOUNDATION SOUTH AFRICA

15 SEPTEMBER 2023

Content

1. What is AI?
2. Diversity
3. Design Alignments
4. Ownership Ambiguities
5. Criminal Activities
6. Guardrails
7. Implications
    - The Silver Lining
8. Concluding remarks

Background: What is AI

- Artificial Intelligence
  - A set of algorithms that mimic human and natural intelligence
  - These Algorithms learn from examples presented as data
  - They learn by finding thousands of parameters that can collectively be used to represent some target variable.

  - As computing power becomes better, these parameters have increased to millions
    - This increases their ability to ‘remember’ examples by large amount of people.
    - This also means that AI can be used to compute large amount in short space of time, faster than humans.

Background: What is AI

- Human mind can process at most 8 dimensions
  - AI models can handle over 1000 dimensions across millions of records
  - This makes it more efficient and accurate in describing said targets faster than humans.

- AI in its current nature is ‘Narrow’
  - This means that it can only perform tasks that it has been trained to generalise.
  - This is different from how human minds perform where we can map our learnings to new tasks with ease.

Just like forex, AI is an embedded product that cannot be used on its own.

Critical pitfalls: Diversity

- **Out of Sample Issues**
  - AI hardcodes knowledge presented by the designer
  - Designer limitations of world understanding dicates encoded knowledge
- **Frequency Bias**
  - Only frequently held beliefs are considered more accurate
  - Frequently occurring beliefs mainly exposed to users
    - Misconceptions can easily be propagated
    - Widely believed misconceptions viewed as accepted facts
      - Eg. Pluto being a planet
  - Herd Mentality
    - As a result of suggesting frequently used items
      - Eg. Most male population loves soccer, so every male should love soccer
- **Systematic Discrimination**
  - Lack of presentation in training sample, leads to being systematically discriminated against.
    - Eg. Females and Africans don't hold executive positions
- **Historical Injustices**
  - These were systematically engraved and humans still made bad judgement.
  - Machines taught on this data, follow same process
    - 35 YO, 5 Years Work Experience, BSc Degree, R20k pm salary, 3 dependency not fit for Home Loan
      - But this is average description of most Black people
      - A white person of same qualities could qualify a first home buyer loan.

Critical pitfalls: Design Misalignments

- **Designers Agendas**
  - Are ALL engagements good engagements?
    - AI can be designed to maximise screen time from users.
    - This can be good for some contents but bad for some
    - Where should Behavioural Economists draw the line?

  - Recommendation Systems Maximise Designers Profit
    - What about users welfare?
      - AI systems can be trained with rewarding systems thus causing addiction to users.
        - Eg, usage of YouTube, TikTok, Twitter.
      - Products/services offered might look appealing but lead to poverty
        - Taking advantage of human's slow processing
          - A danger in impoverished areas.
          - Combo of products user does not need but appears cheaper

Critical pitfalls: Ownership Ambiguities

- Design Responsibilities
  - Who is responsible? Who is accountable for designs?
  - For AI, how far is the engineer responsible for the product given that it is refined on users' details?
  - Responsibility boundaries are blurry

- Design Ownership
  - If users uses Generative AI, who owns the product? The owners of Generative AI or the prompter?
  - But it was improved using community data, shouldn't community have a share?

- If an autonomous car crashes into passengers to protect car owner, who is liable?
  - Car owner vs Car Manufacture?

Critical pitfalls: Criminal Activities

- Adversarial Attacks
  - Users can deliberately change input data to fool the models and make them classify falsely.
    - Eg. Change data input to favour a classification output of the user.

- Falsify Presence using Generative AI
  - With Personalised Generative AI, people styles can be learnt and emulated by AI
    - This can lead to the person whose data was used to be falsely placed in situations
      - Deep Fake
      - Fake emails
      - Improved Spams
      - Improved con chatbots
- Reconstruction of inputs
  - With usage of Generative AI, outputs can be reconstructed outputs.
  - This means: from latent features, original data can potentially be reconstructed.

Pitfalls of AI - Guardrails

- These pitfalls are not unique
  - Other fields and applications had similar pitfalls but guardrails were designed.
- Industrial Guardrails come in forms of
  - Standard Operating Procedures
  - Regulatory Policies,
  - National and International Standards
  - Voluntary Bodies open to public and private participation.
- Most of these are currently available for Data Usage and not Design and Deployment of AI

Pitfalls of AI – Guardrails (2)

- AI has potential to be militarised like Nuclear is.
  - Unlike Nuclear, AI does not require easily traceable materials but requires energy consumption to train models.
  - Should similar policies be investigated and employed to AI development

- In SA, the Customer Protection Act and Competition Commission ensures that clients are fairly treated by companies.
  - Should it be a request that these policies be included as objectives while training the AI models to be used in SA?
  - This should mean that AG is empowered and Upskilled to ensure this is indeed applied

Some remarks: R&D Funding in AI

- Advantages of AI Research in Africa:
  - Inclusion of African context
  - Curb rapidly growing brain drainage
  - Encourage Ethical AI
  - Fully understand the technology:
    - Helps with policy designs
    - Data Governance Policy Designs
    - Avoid importing Digi-Tech like we did Ride-Hailing Platforms
- Most SA companies are embracing Digital Platform businesses
  - Funding AI research can accelerate the growth in these platforms
  - Increase number of employment
- More investment in 'Soft' Engineering' such as App Development, Game Development

Conclusion:

- This is a game changer, which will require Governments, Private Sector and Civil society to work together.

Enkosi. Thank you.
Re a leboga. Siyabonga. Dankie