TECHNOLOGY BRIEFING ON AI IN OUR SOCIAL WORLD

The surge of artificial intelligence (AI) in the social fabric of our lives brings forth intricate dilemmas in human interaction, social norms, and personal identity. As AI influences our communication patterns, decision-making processes, and personal connections, it reshapes our understanding of society, relationships, and self-worth. A rising concern is that unchecked AI could erode authentic human interactions, amplify echo chambers, erode existing systems, and challenge the authenticity of online personas. This necessitates a prompt reconsideration of how we integrate technology into our social structures and the ethical implications of its pervasive influence.

Social Media

- **Algorithmic Content Curation**: Digital platforms like TikTok, Facebook, and YouTube utilize *algorithms to curate content tailored for individual users*. By analyzing the data they gather from likes, shares, comments, and watch time, these algorithms push content that is most likely to be engaging for users. The goal is to maximize user interaction and screen time, which boosts ad revenue for these platforms.
- **Echo Chambers & Filter Bubbles**: This personalized content curation creates what's known as an 'echo chamber' or 'filter bubble' effect. *When users are constantly fed content that aligns with their preferences and beliefs, they become insulated from opposing views*. This has profound implications for societal polarization, particularly amongst youth, as users become more entrenched in their beliefs.
- **Chatbots & User-Facing AI**: Chatbots, like Snapchat's AI chatbot or even TikTok's AI-driven responses, are designed to enhance user interaction by providing instant replies and interactions. They're primarily used for customer service, content suggestions, or novelty features.
- **Content Moderation Challenge**s: Young users, like *Chase Nasca* (who committed suicide at 16 years old after being fed suicide-glorifying content by the algorithm), who show interest in distressing content may be further inundated with similar content, potentially exacerbating their mental health struggles. *While it's unfair to lay the blame solely on TikTok for such tragedies, the platform's algorithm does play a role in what content is frequently shown to users.*
- **Psychological & Social Implication**s:
    - Positive: Many users find a community, get educational content, or experience entertainment on these platforms. For some, it offers an escape, a voice, or even fame.
    - Negative: As seen in the tragic case of Nasca, constant exposure to negative content can deepen feelings of despair. The addictive design of such apps can also lead to increased screen time, fostering isolation from real-world interactions.
- **Sentiment Surrounding AI and Social Medi**a: Popular media such as "The Social Dilemma", "Black Mirror", and "Coded Bias" reflect societal concerns about unchecked technological advancement. They underscore the dangers of AI and algorithms, questioning their ethical implications and long-term impact on mental health and societal structure.
    - *"The Social Dilemma"* particularly delves into the addictive nature of social media and the manipulative tactics platforms employ, driven by AI, to keep users engaged.
    - *"Black Mirror"* presents dystopian futures often rooted in technological advancements and the unanticipated consequences of AI and digital culture.

    - *"Coded Bias"* exposes biases in AI systems, a pertinent issue for social media platforms that promise objectivity but may perpetuate societal prejudices.

Surveillance and Facial Recognition

With the exponential growth in computational power and data collection capabilities, AI-enhanced surveillance systems have pervaded cities globally. One of the most prominent technologies within this domain is facial recognition. As with many technological advancements, facial recognition straddles a line between beneficial applications and potential misuse that can infringe upon civil liberties.

Benefits and Positive Applications

- **Crime Prevention and Detection**: Surveillance cameras equipped with facial recognition can help in the real-time identification of criminals or persons of interest, thus aiding law enforcement.
    - *Case*: In 2019, New Delhi police used facial recognition to identify nearly 3,000 missing children within just four days.
- **Swift Identity Verification**: Facial recognition can streamline processes in airports, offices, and other public places where quick identity verification is beneficial.
    - *Case*: Many international airports have integrated facial recognition into their customs and immigration processes to expedite traveler verification.
- **Lost Persons**: Facial recognition can aid in locating missing persons, especially in crowded places.
    - *Case*: As previously mentioned, New Delhi's initiative in finding missing children.
- **Retail and Banking**: Some businesses use facial recognition for customer personalization or fraud detection/prevention.
    - Case: In China, Alipay's "Smile to Pay" system allows customers to make a purchase simply by smiling at a terminal, offering both convenience and a novel security measure.

Concerns and Negative Implications

- **Mass Surveillance and Privacy Infringement**: Ubiquitous surveillance systems might infringe on individuals' privacy rights, creating a sense of always being watched.
    - *Case*: China's vast network of surveillance cameras, often dubbed the "world's biggest camera surveillance network," has sparked international debate about its impact on privacy and freedoms.
- **Racial and Gender Biases**: Some facial recognition algorithms have shown biases, misidentifying individuals, especially those from minority groups.
    - *Case*: An MIT study found that commercial software from companies like IBM, Microsoft, and Face++ had higher error rates in classifying the gender of darker-skinned and female faces.
- **False Positives**: Misidentification can lead to wrongful accusations or detentions.
    - *Case*: In the U.S., a Black man in Michigan was arrested in 2020 after being misidentified by facial recognition software, marking a cautionary tale about the technology's potential pitfalls. 18-year-old Oursmane Bah was misidentified as an Apple store thief and arrested by the New York
- **Freedom of Expression**: Knowing that one is constantly under surveillance can deter individuals from exercising their rights to free speech and assembly.
    - *Case*: In Hong Kong, there have been reports of protesters tearing down smart lamp posts over fears of state surveillance during the 2019-2020 protests.
- **Data Security**: The vast amount of biometric data collected poses a risk if there's a data breach.
    - *Case*: In 2019, a data breach at a company called Suprema exposed the biometric data of over one million people.

***Note**: Fidutam believes that we should ban facial recognition and public surveillance systems that classify individuals on factors of race, gender, religion, and other sociocultural identities. Most, if not all use cases of such systems are predatory and often discriminatory if used as a basis for prediction.*

Extended Reality

The digital age has witnessed the evolution of interactions from tangible, real-world settings to virtual realms, drastically changing the dynamics of communication and experiences. Augmented Reality (AR), Virtual Reality (VR), Mixed Reality (MR), and Extended Reality (XR) have become the frontier of this transition, with AI serving as the backbone that elevates the virtual experience.

Definitions:

- *AR* (*Augmented Reality*): Superimposes digital information or graphics onto the user's real-world view.
- *VR* (*Virtual Reality*): Transports users to a fully digital environment, detaching them from their physical surroundings.
- *MR* (*Mixed Reality*): A fusion of AR and VR, wherein digital and physical objects co-exist and interact in real time.
- *XR* (*Extended Reality*): An umbrella term encapsulating AR, VR, and MR, representing all forms of computer-altered reality.

Benefits and Positive Applications

- **Enhanced Learning and Training**: AR and VR can provide immersive experiences for learners, from medical training to history lessons.
    - *Example*: Oculus, owned by Meta, offers VR educational applications that transport students to space or historic moments for an experiential form of learning.
- **Healthcare**: AI-driven VR therapies are being researched for pain management, PTSD, and cognitive rehabilitation.
    - *Example*: Patients with PTSD have undergone VR-assisted exposure therapy, wherein AI fine-tunes the virtual scenarios based on real-time physiological feedback.
- **Inclusive Digital Social Interactions**: VR allows for inclusive interactions, enabling people with disabilities to engage in ways that might be challenging in the physical world.
    - *Example*: VR chat rooms where individuals with mobility issues can "walk" and "talk" as avatars.
- **Advanced User Experience**: Devices like Apple's anticipated Vision Pro headset for 2024 are expected to use AI-powered eye-tracking to personalize the AR experience by understanding user focus and intention.

Concerns and Negative Implications

- **Impersonation and Authenticity Issues**: As AI-generated avatars become increasingly realistic, there's potential for misuse in terms of impersonating real individuals.
    - *Example*: Deepfakes (typically used to refer to fake or edited AI-generated media) in VR environments can lead to situations where users might believe they're interacting with real people when they're AI constructs.
- **Mental Health Concerns**: Prolonged exposure to VR can lead to detachment from the real world, potentially exacerbating conditions like depression or triggering disassociative experiences. *With citizens aged 25 experiencing the strongest ties to digital platforms, young demographics are especially at risk.*
- **Data Security and Privacy**: The integration of biometrics and neurological feedback in XR devices presents a novel frontier for potential data breaches.
    - *Example*: If companies like Apple, integrating eye-tracking and neurological sensors into their XR devices, were to suffer a data breach, it could expose incredibly intimate data about users' reactions and even thoughts.

        - There's also a chance that *unethical actors could monetize or illegally source this data*.
- **Economic Disparities**: As AR and VR tech become integral to sectors like education, those unable to afford these technologies risk being left behind.
- **Manipulated Realities**: AI-enhanced AR could alter one's perception of reality, intentionally or otherwise. Misinformation and propaganda can be embedded into these realities.
    - *Example*: An AR application that overlays false historical information on real-world monuments.

Interconnected Systems

The fusion of Artificial Intelligence (AI) with the Internet of Things (IoT) has reshaped our modern world, enabling intelligence-driven devices and systems in homes, industries, and cities. The increasingly localized nature of this AI, often referred to as TinyML, further amplifies the capabilities of these devices while presenting new challenges.

Definitions:

- IoT (Internet of Things): A network of interconnected physical devices that communicate and exchange data without human intervention. This includes everything from smartphones to smart fridges, smartwatches, and even smart cities.
- Local AI/TinyML: Tiny Machine Learning, or TinyML, refers to AI models optimized to run on low-power and low-resource devices. These are "edge" devices that can compute locally without needing to constantly relay information to the cloud, increasing security and processing speed but reducing capability.

Benefits and Positive Applications

- **Efficiency and Predictive Maintenance**: Sensors in IoT devices, powered by AI, can predict when machinery will fail or when a system needs servicing. This leads to cost savings and prolonged equipment life.
    - *Example*: In industries, smart sensors detect anomalies in machinery, allowing for timely maintenance and preventing production downtime.
- **Enhanced User Experience**: AI learns from user behaviors and preferences to automate tasks in smart homes. Lights, temperature, and entertainment systems can be adjusted in real time based on user moods or schedules.
- **Energy Conservation**: Smart cities employ AI to regulate traffic, control lights, and manage waste, leading to a decrease in energy consumption and emissions.
- **Decentralized Processing**: With TinyML, data processing happens at the device level, reducing the need for constant internet connectivity and reducing data transmission costs.

Concerns and Negative Implications

- **Security Vulnerabilities**: The interconnectivity of IoT devices makes them susceptible to hacks. A breach in one device can potentially compromise the entire connected system.
    - *Example*: A hacker gaining control over a smart thermostat could manipulate other connected devices in a home, from security systems to kitchen appliances.
- **Data Privacy**: Even with localized AI, there's potential for misuse or unauthorized access to personal data.
    - *Example*: If a smart fridge tracks eating habits, this data, if accessed, could be used by insurance companies to adjust premiums based on perceived health risks.
- **Dependency and System Failures**: Over-reliance on smart systems might lead to challenges if they fail. People may become less self-reliant or adept at manual tasks. Furthermore, continued digital migration will exclude rural stakeholders or populations in lower-infrastructure areas from accessing essential services, tools, or utilities.
- **Economic Implications**: The shift to smart industries could result in job losses in traditional sectors, leading to economic disparities and the need for workforce upskilling.

- **AI-Facilitated Scams and Vulnerability of the Elderly**: AI's capability to mimic voices has turned into a weapon for scammers. Using data sourced from various interconnected sources, like social media or public videos, AI can be trained to imitate a person's voice.
    - *This technique is particularly concerning for the elderly*, who may not be as tech-savvy and might find it harder to discern between genuine and AI-generated calls. The AARP has highlighted that while younger individuals might encounter fraud more frequently, it's the older population that stands to lose more, both emotionally and financially.
        - *Example*: The case of Jennifer Destefano, whose AI-generated daughter's voice led her to believe her child was in distress, showcases the emotional turmoil and potential financial threats posed by these scams.

Algorithmic Systems

Algorithms are being applied in all areas, from recidivism prediction to research to creating online content, making it almost impossible to cover the entire–and continually expanding–panorama of AI's integration into social systems. Outlined below are five important social issue areas to consider.

A. **Housing and Real Estate**
    a. Positives
        i. *Trend Analysis*: AI is a game-changer for real estate professionals in predicting housing market trends. It can analyze vast amounts of data quickly, identifying emerging hotspots or detecting early signs of a market downturn. For example, platforms like Zillow use predictive analytics to give homeowners an idea of how much their property might be worth in the future.
        ii. *Optimized Pricing*: AI helps sellers and realtors determine the most competitive price for properties based on similar listings and historical data. This can lead to quicker sales and potentially higher profits.
        iii. *Streamlined Property Discovery*: For buyers, AI-driven platforms can tailor property suggestions based on individual preferences, such as proximity to schools or workplaces, thus providing a personalized experience.
    b. Negatives
        i. *Digital Redlining*: AI models trained on historical data might perpetuate longstanding biases. If a particular community was historically undervalued, AI might continue this trend, leading to unfair practices like digital redlining.

B. **Employment**
    a. Positives
        i. *Efficiency in Recruitment*: AI tools, such as Pymetrics, analyze resumes and applications much quicker than human recruiters, speeding up the hiring process.
        ii. *Predictive Performance Analysis*: Predictive algorithms can forecast which candidates are more likely to be successful in a role based on their skills, experiences, and other factors.
        iii. New AI applications are emerging to facilitate salary negotiations and reduce demographic bias by having faceless, (seemingly) neutral systems engage employees on sensitive matters.
    b. Negatives
        i. *Bias Amplification*: If not properly calibrated, AI recruitment tools can perpetuate existing biases, leading to discriminatory hiring practices. For example, Amazon once scrapped its AI recruiting tool after finding it biased against female candidates.

ii. *Impersonal Process*: Over-reliance on AI in recruitment may lead to impersonal hiring processes, missing out on potential candidates who might be a cultural fit but don't match the AI's criteria.

C. **Healthcare**

a. Positives

i. *Diagnosis Prediction*: Companies like PathAI are using machine learning to aid pathologists in diagnosing diseases, often with greater accuracy.

ii. *Personalized Treatment*: AI can analyze a patient's data to suggest personalized treatment plans, maximizing effectiveness and reducing side effects.

iii. *Operational Efficiency*: Hospitals and clinics utilize AI to optimize appointment scheduling, manage patient flow, and predict patient needs.

b. Negatives

i. *Data Privacy Concerns*: With AI's reliance on vast amounts of patient data, there are growing concerns about data breaches and unauthorized access.

ii. *Malfunctions*:

1. Bias in Treatment Recommendations: A 2019 study published in Science uncovered that an algorithm widely used in U.S. hospitals to allocate health care to patients has been systematically discriminating against Black patients. *The algorithm assigned white patients the same level of risk as Black patients despite the Black patients being notably sicker*. This led to a significant disparity in the healthcare programs to which patients were referred. The issue stemmed from the algorithm's design which used healthcare costs as a proxy for health needs. Historically, Black patients incurred lower costs, which caused the algorithm to underestimate their health needs.
2. Misdiagnoses Due to Dataset Limitations: Some diagnostic tools were trained primarily on data from individuals of European descent. When applied to individuals of African, Asian, or other descent, the tools might not perform with the same level of accuracy. For example, certain skin conditions might manifest differently across skin types, and *if a dermatology AI was predominantly trained on lighter skin, it might misdiagnose conditions on darker skin*.
3. False Positives and Negatives in Imaging: Deep learning models used for medical imaging sometimes produce false positives or miss critical anomalies, especially when exposed to cases not well represented in their training data. *An AI algorithm designed to predict malignant tumors was incorrectly flagging them because it had correlated the presence of a ruler in a medical image to a tumor being cancerous* (a coincidental trend that it had found in its training data).
4. Over-reliance on Wearables: Devices like heart rate monitors and sleep trackers can sometimes produce inaccurate data, leading users to seek unnecessary medical treatment or overlook real health issues. For instance, while devices like the *Apple Watch have successfully detected atrial fibrillation in some users, they might also produce false positives*, leading to unwarranted anxiety and medical visits.

D. **Education**
   a. Positives
      i. *Adaptive Learning*: AI tools can customize learning experiences based on a student's pace and learning style. Platforms like DreamBox offer adaptive math programs that adjust to the learner in real time.
      ii. *Personalized Learning*: Systems like ChatGPT can be used to provide personalized and effective learning systems to students. For example, in Fidutam's educational study, a GPT-4 (ChatGPT's AI model)-based curriculum generation platform was developed. With it, over 2,400 underserved students were able to learn math and science topics without access to textbooks or specially-trained teachers, including those with learning disabilities.
      iii. *Enhanced supplemental learning*: Tools like Khan Academy's Khanmigo and Quizlet's Q-Chat provide customizable and automated tutoring services and interactive learning experiences such as speaking to digital twins of historical figures (though there are concerns about the misrepresentation of these influential people, particularly nonliterate ones such as Harriet Tubman).
   b. Negatives
      i. *Unfair Assessment*: As seen with the case of Louise Stivers, the use of AI like Turnitin to detect not just plagiarism but also whether an assignment is AI-generated can sometimes flag genuine student work, causing undue stress and challenges.
      ii. *Cheating Epidemic*: AI tools like ChatGPT are leading to concerns about academic dishonesty. Half of college students might already be using such AI tools to cheat, eroding the very foundation of educational assessments.
      iii. *Policy Challenges*: Institutions face the uphill task of updating policies to encompass new forms of cheating and to understand the implications of AI tools in academic environments.

E. **Criminal Justice**

With the rise of technology and data analytics, algorithms have been introduced into the justice system to predict an individual's likelihood of committing a future crime or reoffending. These predictive tools aim to aid in decisions related to bail, sentencing, and parole.

   a. Examples
      i. *COMPAS* (*Correctional Offender Management Profiling for Alternative Sanctions*): One of the most widely known risk assessment tools, COMPAS, is used to predict an offender's risk of recidivism. It assesses risks based on answers to 137 questions, covering criminal and parole history, age, and gender, among others.
      ii. *Public Safety Assessment* (*PSA*): Developed by the Laura and John Arnold Foundation, the PSA uses nine factors from an individual's criminal history to predict the likelihood of a new crime or missing a court appearance.
   b. Positives:
      i. *Consistency Across Decisions*: Algorithms can provide a standardized metric across different courts or jurisdictions, ensuring a more uniform approach to risk assessment.
      ii. *Potential to Reduce Incarceration Rates*: When used correctly, these tools could identify low-risk individuals who don't need to be detained, reducing jail populations.

c. Negatives:
    i. *Inherent Bias in Training Data*: Algorithms are trained on historical data, which can perpetuate systemic biases. For example, if past judicial decisions were racially biased, the algorithm could learn and continue this bias.
    ii. *Transparency and Interpretability Issues*: Tools like COMPAS have been criticized for being "black boxes," where the decision-making process is not transparent. Judges, defendants, and the public often don't know how the risk scores are derived.
    iii. *Inaccuracy in Predictions*: A notable investigation by ProPublica found that the COMPAS system was only accurate in its recidivism predictions around 65% of the time. Additionally, the study found that the system falsely flagged Black defendants as future criminals at almost twice the rate as white defendants.
    iv. *Overpolicing*: Software to predict or monitor criminal tend to be piloted in areas that already receive a disproportionate amount of policing. For example, a Chicago 13-year-old was shot and killed by police while unarmed after ShotSpotter (a gunshot detection software) pinged his location.
    v. *Ethical Implications*: There's a fundamental ethical concern about determining an individual's future actions based on historical data or broad categorizations. It challenges the principle that justice should be individualized.
d. Stakeholder Effects
    i. *Defendants*: Those subjected to these algorithms may face longer sentences, higher bail amounts, or stricter parole conditions based on potentially flawed or biased predictions.
    ii. *Judicial System*: Judges and other court officials, despite having discretion on whether to use or weigh the results of profiling tools, may come to over-rely on these tools at the expense of individual judgment or become too trusting of their accuracy.
    iii. *Society at Large*: While there's a potential for safer communities if high-risk individuals are accurately identified, there's also a risk of perpetuating biases and injustices on a systemic level, and unfairly suppressing or surveilling already vulnerable populations.
e. Note: The use of predictive algorithms in justice mirrors a broader trend in society where data-driven tools are being employed in critical decision-making processes, from hiring to loan approvals. However, the stakes in criminal justice are particularly high, as decisions can deprive individuals of their freedom and significantly impact their prospects.

As AI becomes deeply embedded in our social interactions, clear guidelines and oversight are crucial. Engaging experts from tech, sociology, and ethics, as well as stakeholders (*especially* youth and underserved populations), can help create a balanced approach. The goal is straightforward: leverage AI's potential to enhance communication while avoiding pitfalls that could harm societal bonds. The future of our digital society depends on getting this balance right.