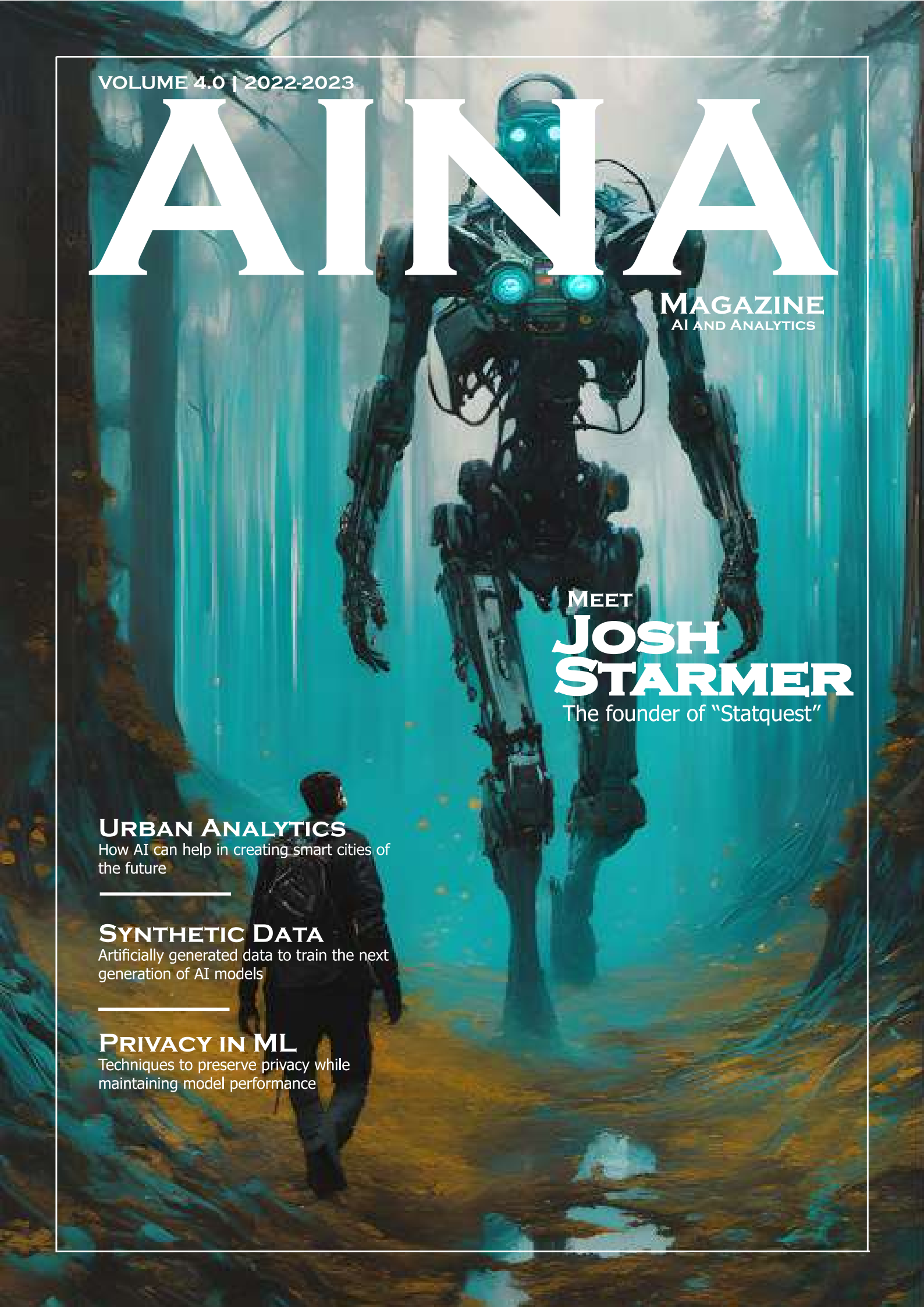
VOLUME 4.0 | 2022-2023

AINA

Magazine

AI and Analytics



Meet Josh Starmer

The founder of "Statquest"

Urban Analytics

How AI can help in creating smart cities of the future

Synthetic Data

Artificially generated data to train the next generation of AI models

Privacy in ML

Techniques to preserve privacy while maintaining model performance

EDITOR IN CHIEF:
Writtabratta Bhattacharya

LEAD DESIGNER:
Utkarsh Sinha

CREATIVE HEAD:
B.N.K. Goutham

PR HEAD:
Raahul N

FEATURES EDITOR:
Aravintha Kannan

SUB EDITOR:
Parijat Pushpam

From The Team

Artificial Intelligence has firmly established itself as a transformative force, reshaping industries by automating mundane tasks and driving ground-breaking research. Its omnipresence spans healthcare, finance, automotive, entertainment, and beyond. It feels like just yesterday when Stable Diffusion and GPT-3 sent shockwaves through the AIML world, and now Large Language Models (LLMs) are gaining traction in sectors like Healthcare and Finance. Researchers are taking bold strides to turn once-unthinkable advances into reality, and what seemed like dreams a decade ago is now at our fingertips.

We say, "Keep Dreaming!" The best ideas often emerge from the depths of our imagination, and humans have strived to recreate that imaginative prowess artificially. Who knows what breakthroughs lie ahead? This leads us down the path of Generative AI, a term that resonates with both the youngest and most seasoned professionals. The quest to imbue the systems we create with human-like imaginative capacities unites us all. The world is curious, and people's thirst for knowledge propels us to great heights. While the development we witness sparks mixed responses, with notable figures like Geoffrey Hinton and Elon Musk expressing concerns about the potential to replace or manipulate humans, other AIML leaders like Yann LeCun and Sam Altman remain optimistic about this exponential progress in AI.

As we eagerly await the future, let's take a moment to reflect on the developments of recent years. 'AINA – Artificial Intelligence and Analytics,' India's pioneering student-driven analytics magazine, embarked on its journey in 2020 with the aim of promoting awareness of data analytics in the industry and the advancements in AIML. In this fourth edition of AINA, we are thrilled to present the latest global trends and developments in these fields, showcasing the diverse applications of AI and analytics across various domains. Our magazine's design concept has been carefully crafted to align with its theme and capture the dynamic, creative spirit of the entire PGDBA community.

We extend our heartfelt gratitude to the chairpersons, directors, deans, and faculty of ISI Kolkata, IIM Calcutta, and IIT Kharagpur for their unwavering guidance and support. We are immensely thankful to Dr. Joshua Starmer, the founder of Statquest, and Mr. Udai Shankar, Director of Data Science at Providence, for sharing their invaluable insights and perspectives through interview sessions that will provide our readers with a broader outlook. We'd also like to acknowledge the unwavering support, guidance, and suggestions from our senior magazine team members from AINA 3.0, who contributed significantly to our final product.

Our magazine is the result of blending lessons from our educators, the support of friends and family, and mentorship with a dash of imagination and a cup of honest effort. We present this magazine to you with great pride and anticipation. Happy reading!

From the Chairperson's Desk

Post Graduate Diploma in Business Analytics (PGDBA) is a full time, residential program in business analytics offered jointly by three esteemed institutes: Indian Institute of Management (IIM) Calcutta, Indian Statistical Institute (ISI), Kolkata, and Indian Institute of Technology (IIT), Kharagpur. Naturally, the uniqueness of this program emanates from the amalgamation of business, statistics, and technology. The program conducts an extremely competitive selection process and maintains a good mix of experienced and fresh graduates to build an impressive batch profile over the years. More importantly, unlike any other program, an industry internship of six months stands this program out from the rest of its cohort as it makes the participants completely industry ready. The rigorous and demanding training imparted by the three institutes through theory, case studies and business simulations prepare the students for the world of business analytics. In addition, students of PGDBA organize a pioneer, tri-institute flagship analytics summit of India, Trilytics, that features an analytics case competition at global level. Such a rich exposure in analytics helps students build the ability to think critically and find data driven solutions to management problems using appropriate tools from the field of data science. I am sure that this unique tri-institute PGDBA program will grow more and more in future to cater the increasing demand of data scientists, globally.

Chairperson, PGDBA

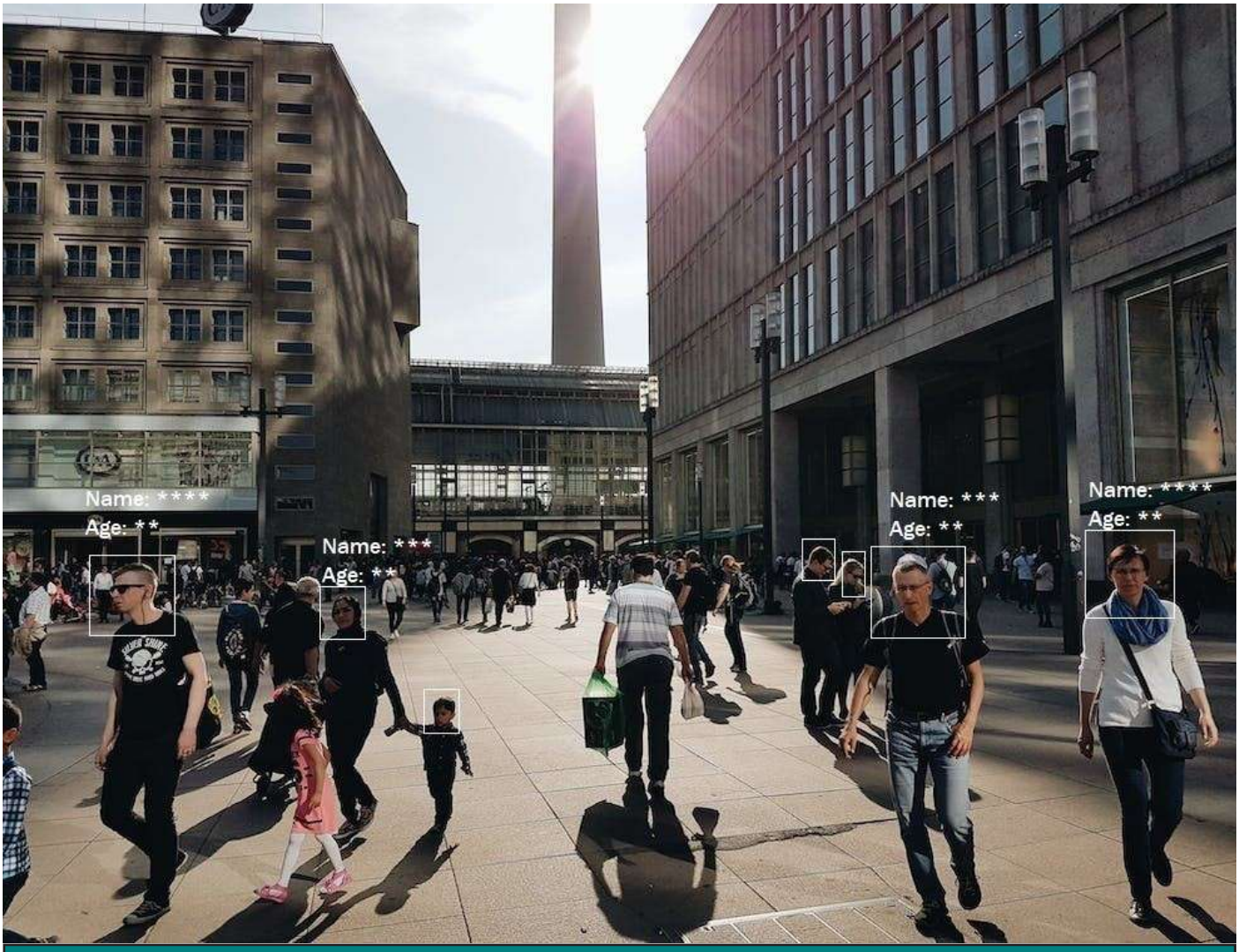
FEATURES

EXPERTISE

Josh Starmer
Founder and CEO - StatQuest
21

Udai Shankar
Director - Data Science, Providence
35

QUICK READS



Algorithms to Allies

Shaping The Future of Policing

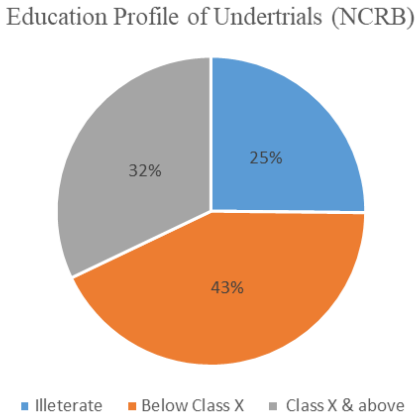
B.N.K. Goutham

Timeless Relevance of Law and Order

Law and order is a concept that has been around since time immemorial, but it has never been more relevant than today. As societies grapple with complex challenges, the need for a structured framework to maintain peace and stability becomes more apparent than ever. It came into existence from the assumption that humans possess inherent inclinations toward self-interest and violence. Beyond preserving societal balance, it presents a cautionary vision of a dystopian future, underscoring the criticality of maintaining a just and regulated system. Imagine a society where law and order don't exist in society. Where people can do whatever, they want, without any consequences. Where violence, chaos, and anarchy reign supreme. Where survival is the only goal and morality is a luxury. This is a world where fear, greed, and selfishness are the only laws. Where justice, peace, and harmony are only dreams. These societies still exist in our modern world, where people have no rights and are living in constant fear.

Unraveling The Maze: Shortcomings

While extreme cases of lawlessness and chaos may be more commonly associated with monarchies and anarchies, it is important to recognize that even in democracies like India, there exist certain shortcomings in maintaining law and order. One small example of this aspect is the bottlenecks faced by Indian Police and prisons. The police force acts as the cornerstone of this foundation. However, the police force faces several challenges, particularly in a middle-income country like India, where resource and budget allocation are key issues. According to a report from the National Crime Record Bureau (NCRB), there has been a significant increase (approximately 15%) in the number of undertrials in Indian prisons. This can be attributed to inefficient resource allocation, accumulation of pending cases, and lack of awareness about one's rights.



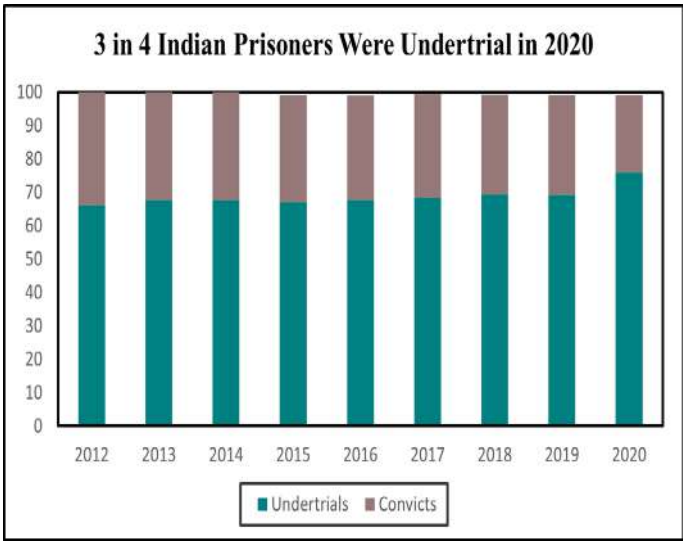
More than 50% undertrials are literate and more than 32% studied class X or above

specific to certain types of crimes, such as the IT Act, POCSO, and Wildlife Protection Act. With such a vast amount of legal data, it is humanly impossible to be well-versed in all aspects of the law. While individuals are typically knowledgeable about crimes that made headlines, there is a lack of awareness regarding other criminal incidents that occur. Lawyers and lawmakers also often find navigating through numerous legal codes, specific laws, and past judgments tiresome. An additional problem is the lack of awareness regarding laws and the legality of issues. Surprisingly, a significant proportion of prisoners (around 52%) are literate, having received education up to at least class X. However, many of them remain unaware of their Fundamental Rights, leading to unfair treatment.

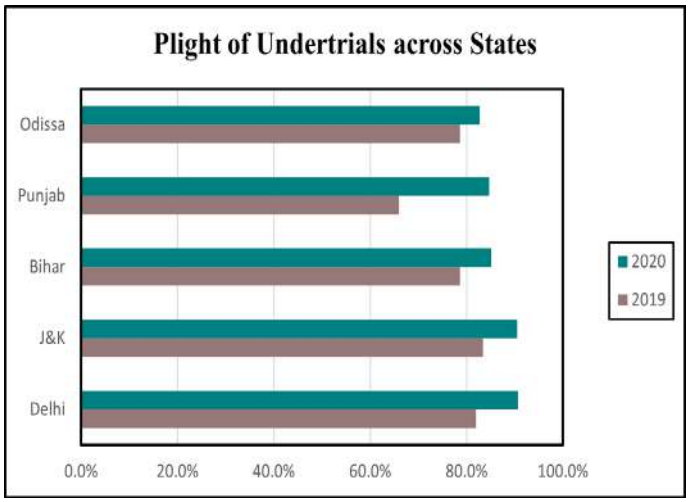
"Society cannot exist unless a controlling power upon will and appetite be placed somewhere, and the less of it there is within, the more there must be without."

- Edmund Burke

Given these budgetary constraints, enhancing the efficiency of the law enforcement mechanism becomes crucial to maintain law and order. Another example of this aspect is the legal illiteracy of the Indian Population and the pile of legal Documentation in India. Currently, India has four main legal codes: the Indian Penal Code, the Code of Criminal Procedure, the Evidence Act, and the Code of Civil Procedure. In addition, there are numerous other laws



Constant rise in the % share of undertrials in Indian Prisons overtime years

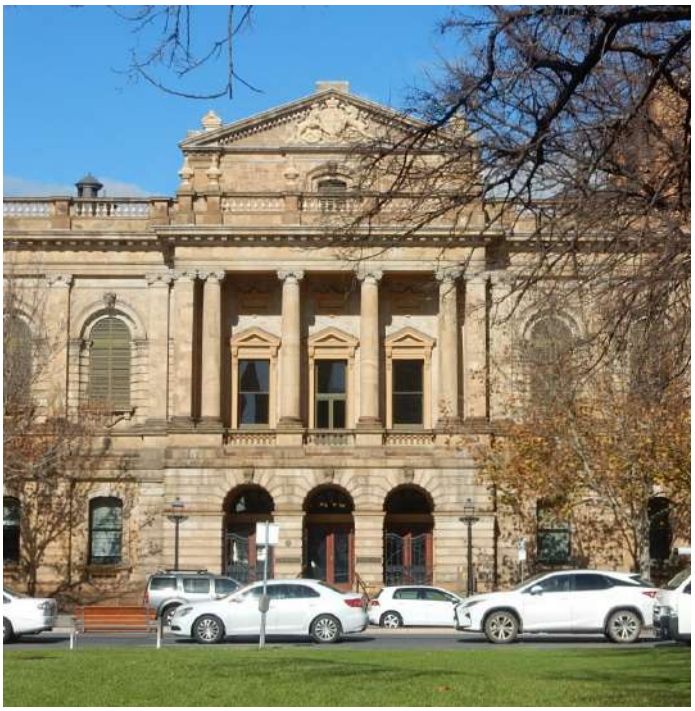


In the year 2020, close to 91% Prisoners in Delhi are undertrials

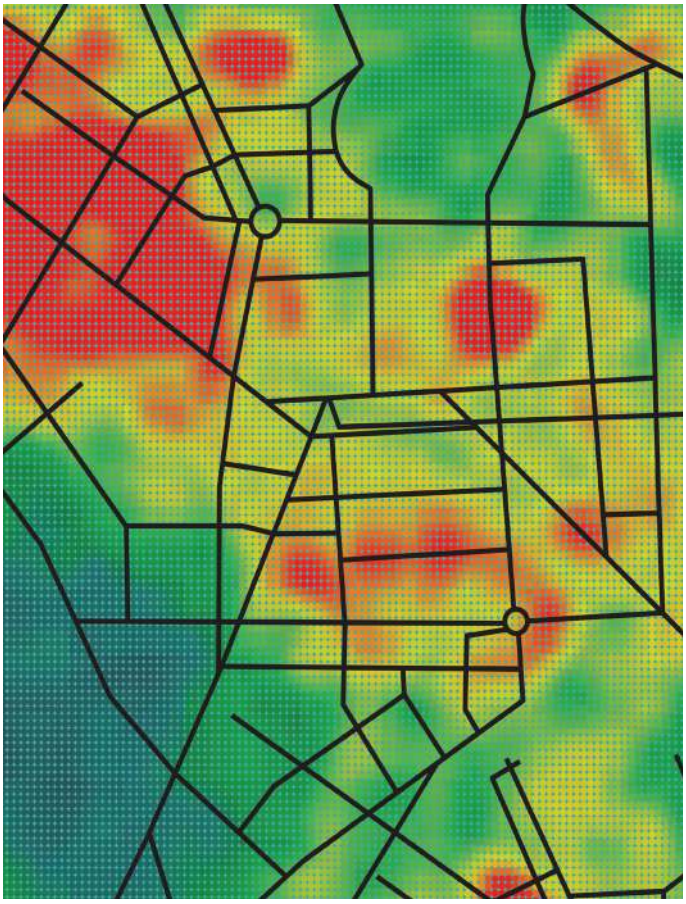
How is AI helping us?

With the recent revolution in AI, its applications are creeping into every field irrespective. This is the case with law and order too. For example, the legal chatbots. chatbots utilizing Natural Language Processing (NLP) and Entity Recognition-based models can offer preliminary legal advice and guidance to individuals. For instance, they can provide information about basic legal procedures, explain legal terms, or direct users to appropriate resources for further assistance. Legalbot is one such example of this. The applications of NLP and ER models extend beyond chatbots; they are also employed in legal research and contract review. NLP-powered legal research aids in identifying relevant laws, statutes, and precedents, while AI automates the review of contracts, reducing the likelihood of errors that may occur during the manual review. Apart from these, the adoption of AI/ML in law enforcement is rapidly advancing, particularly in areas involving large amounts of routine administrative work.

Another significant example of AI and ML in law enforcement is crime prediction. Since the 18th century, efforts have been made to identify patterns in crime using basic mapping techniques. Crime locations were manually plotted on maps to analyze crime concentration and distribution. In recent years, Artificial Intelligence has advanced to the point where police can even predict crimes before they occur.



Ailiria is an Australian company offering legal advice through chatbots

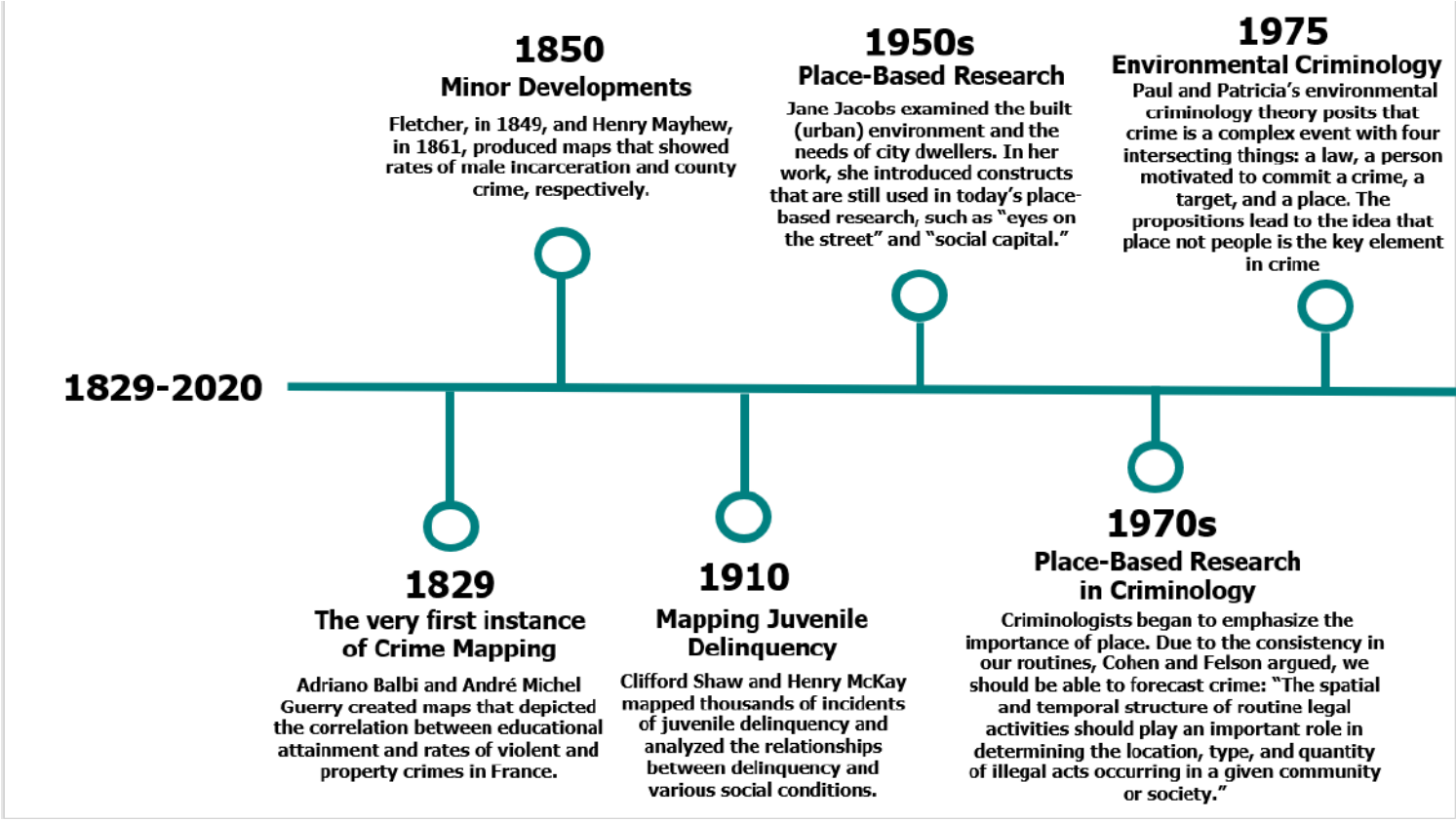


Crime Mapping in the very initial years.

image source: National Institute of Justice

Unlocking Tomorrow's Secrets: Can Data Analytics Predict Crime?

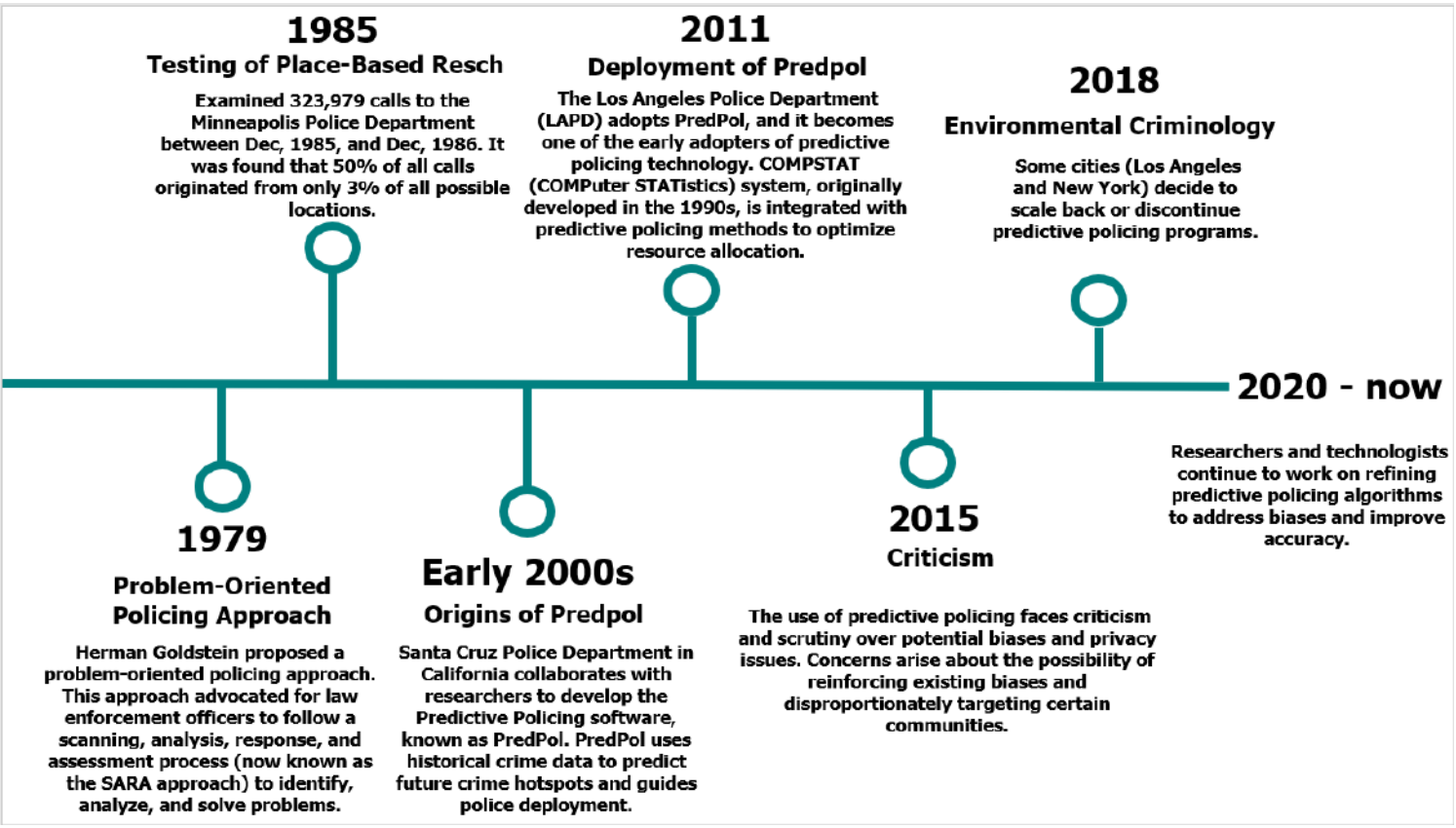
A persistent regret following the occurrence of a crime is that it could have been prevented if we had been aware of it sooner. Even in societies with highly efficient law enforcement systems, this lamentation remains prevalent. However, with the growing scope of predictive analytics, we are edging closer to realizing this dream of predicting crimes before they happen. How is this possible? The simple answer is data. Crime surveillance and data collection have always been integral to policing, whether in societies with highly efficient law and order or in more chaotic environments. Since the 1970s, various types of criminal cases, their locations, times, and dates have been recorded. This constant record-keeping of crime has revolutionized the application of technology in policing. **But how does this data actually help in policing?** By leveraging past crime data, hotspots can be identified based on the type of crime. These hotspots indicate areas with a high likelihood of different crimes.



The Pathways: Types of Predictive Policing

Predictive policing can be broadly categorized into two types: location-based and person-based. As mentioned earlier, location-based predictive policing focuses on identifying high-risk areas or hotspots where criminal activities are more likely to occur. High-risk zones are identified using historical data. More resources and patrolling will be allocated to these specific locations. All this started in the 1970s when the significance of place in the occurrence of crime was realized. The environment criminology theory has put place as the fourth dimension of crime (after person, target, and law). Between December 1985 and December 1986, a total of 323,979 calls were examined by Minneapolis Police Department. Through analyzing real addresses and intersections, the research team discovered that a mere 3% of all potential locations accounted for 50% of all reported calls. From here, research on location-based predictive models has started to shape up. With the increasing computation, model building has become much more simplified. By the year 2005, Los Angeles Police Department has implemented the first model (Predpol) of predictive policing. Person-based predictive policing involves focusing on identifying the persons who are at high risk of committing crimes or becoming victims. Data for this is derived from various sources like criminal records i.e., individual's past arrests, convictions, demographic data, geographic data, social network data, probation data, etc.

The Los Angeles Police Department implemented the first model of predictive policing in 2005



The Trailblazing Journey

Predpol is the earliest predictive analytics algorithm used in predictive policing. It is based on a decade of detailed academic research into the causes of crime pattern formation. Thus, research was successful in linking the key aspects of offender's behavior to mathematical structure that is used for predicting the pattern in the evolution of crime from day to day. In Predpol, only the historical crime data is considered for model building.

Five data points are used from each incident used for predictive analytics.

Incident identifier: Unique ID given to every incident by department (Acts as primary key).

Crime Type: Crimes are segregated into various categories. This category is taken as one of the input parameters.

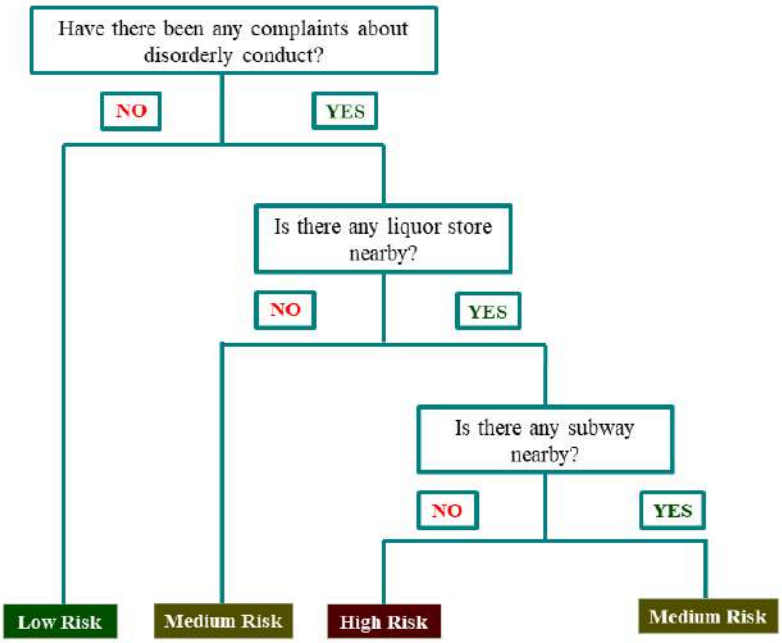
Location of Crime: Latitude and longitude are used for better accuracy.

Time stamps: The exact time of occurrence of crime is not known in all cases. To counter this, we use time stamps i.e., the beginning date, time, and ending date and time.

Record modified date or time of incident: In case of any changes in timestamps, this is used. Predpol's algorithm uses these as input points, and divides the whole space into small grids, with a size of 500 square feet. The crime is predicted in these boxes/clusters. Since different kinds of crimes are considered, respectively required police force and resources are calibrated.

Building upon the foundation of PredPol, a more advanced technique called Risk Terrain Modeling (RTM) has been developed. While PredPol uses only uses past data of crimes as a parameter to predict the crime, Risk Terrain Modelling uses spatial analysis considering various environmental factors like socio-demographic data, that contribute to the occurrence of crime, as a parameter to predict the occurrence of crime.

In the context of Risk Terrain Modelling (RTM), the identification of environmental factors extends to highly specific elements, including the proximity to public transport, local weather conditions, and other precise contextual details. These factors are carefully considered to gain a nuanced understanding of the spatial dynamics that may influence crime occurrences. By incorporating such granular information, RTM enhances the accuracy and effectiveness of predictive policing strategies, allowing law enforcement agen-



Basic Structure of Hunchlab, which works on a tree based mechanism for predictive policing

cies to tailor interventions and allocate resources in a more targeted manner. Different statistical analysis is conducted to identify the relationship between these environmental factors and crime events. From these, risk maps are created that highlight area with high risk. The idea of considering environmental factors in the occurrence of crime can be dated back to the article written by Dr. Joel Caplan and Dr. Leslie W Kennedy. This model is adopted by police departments in Philadelphia, New York, Dutch, and London.

HunchLab represents a significant advancement beyond the Risk Terrain Model, incorporating a sophisticated tree-based model that assigns risk levels to each grid cell. This cutting-edge approach greatly enhances the accuracy and precision of risk assessment, enabling more effective prediction and analysis

Catching the Future: The Prospects and Pitfalls

Indeed, like any technology or approach, predictive policing comes with its own set of drawbacks and challenges. The very first one starts with the data itself. Predictive policing algorithms function best under ideal conditions of fair and unbiased data. Unfortunately, this is not always the case in the real world. For example, the Chicago Police Department (CPD) has a history of corrupted and biased data against marginalized communities.

Utilizing such biased data can result in algorithms that perpetuate biases and exacerbate existing issues, worsening the situation instead of improving it. It is essential to address existing biases within law enforcement before implementing predictive policing technologies, as the accessibility of transformative technologies like machine learning and artificial intelligence can reshape the landscape of law and order.

Lack of transparency is another major issue observed in law enforcement practices and the development of predictive algorithm-based software. This hinders public scrutiny of data usage and procedures, as the inner workings of these algorithms are often undisclosed to the public.

For instance, the Los Angeles Police Department's LASER program, a prediction algorithm-based system, was discontinued in 2019 due to an internal audit report revealing irregularities within the program.

Similarly, the Chicago Police Department had to abandon its predictive software-based program due to severe criticism.

All these shortcomings raise a fundamental question: Can predictive policing techniques be effectively adopted in India? Will their implementation help enhance the law enforcement system, or will it result in an undue concentration of power in the hands of the state, potentially leading to further challenges and unrest?

Data-driven policing programs can reinforce harmful patterns, fueling the over-policing of Black and brown communities

The Indian Context

AI applications have entered the Indian Police force relatively recently (since 2019), but they have already made significant strides. For example, the Crime Mapping Analytics and Predictive System of Delhi collects data from ISRO's satellites every three minutes and uses the Dial 100 helpline to identify crime hotspots. The Hyderabad Police Department has started using person-based predictive analytics algorithms, utilizing data from the Integrated People Information Hub, which includes sensitive details like biometrics, bank information, passport details, and Aadhar card information.

Considering the continuous threats from internal and external forces and the existing issue of underfunding, there is a necessity for the Indian police force to incorporate predictive algorithm techniques into their daily operations. However, it is crucial to exercise caution before implementing large-scale machine learning-based algorithms.

Challenges such as under-reporting of crime in certain states like Uttar Pradesh and Bihar can lead to the generation of flawed data. There is also the potential for bias against minority and underprivileged sections of society, given that a significant portion of the existing undertrials belongs to these communities. Transparency concerns arise, particularly when sensitive information like fingerprints and retinas are utilized.

It is important to note that merely enhancing the efficiency of the policing system does not provide a comprehensive solution. It can create bottlenecks in other areas, such as the judiciary, and unbalanced power dynamics within the three pillars of democracy.

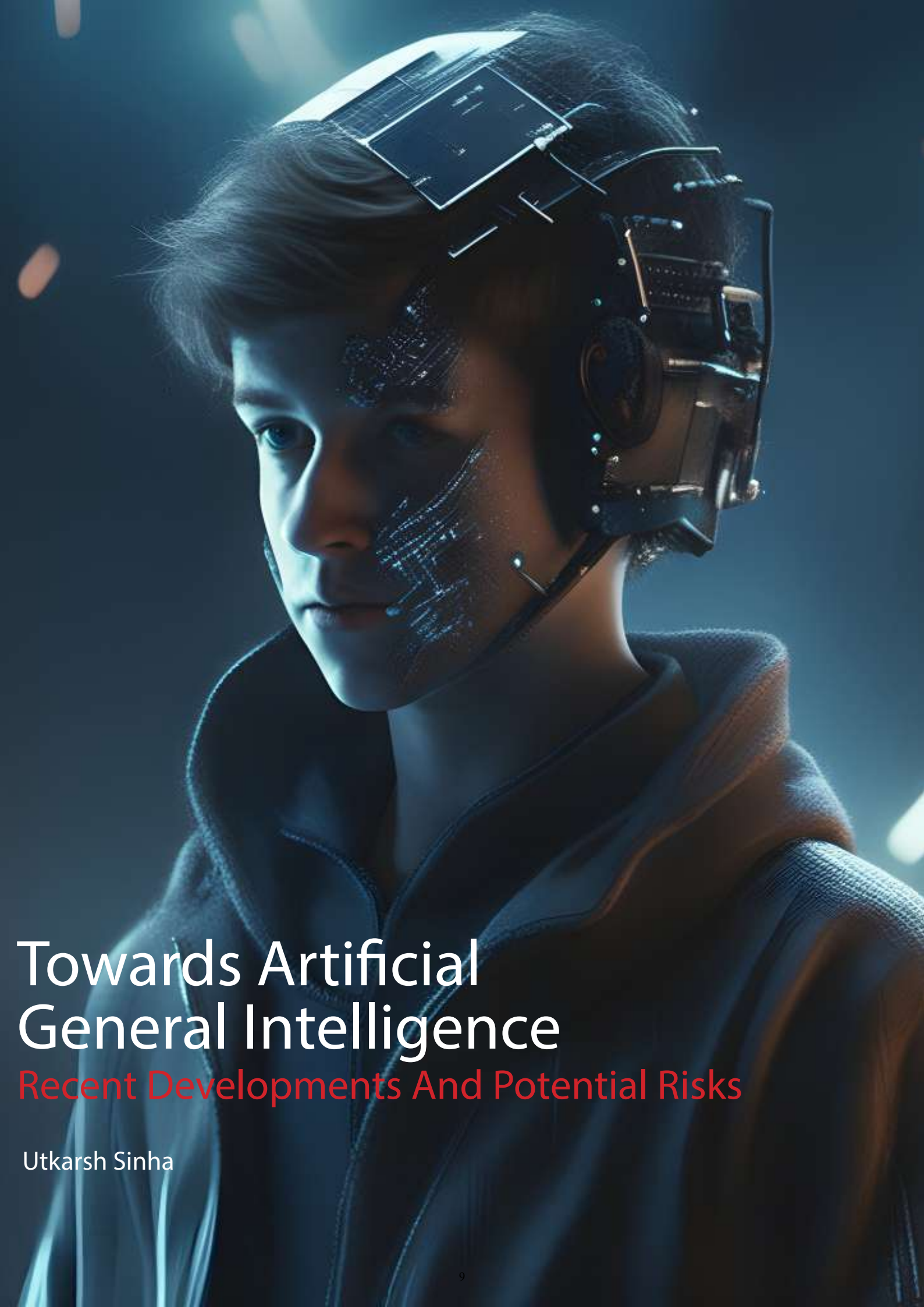
Therefore, while considering the adoption of predictive policing techniques in India, careful evaluation and appropriate safeguards must be implemented to address these concerns effectively. It is crucial to strike a balance between enhancing law enforcement capabilities and preserving the democratic principles and rights of individuals.

"Surveillance is a necessary evil in the face of evolving threats, but it must always be accompanied by strong oversight and safeguards to protect individual liberties"
- Edward Snowden

.

Conclusion

The adoption of AI and ML in law enforcement has the potential to enhance efficiency, optimize resource allocation, and aid in crime prevention. However, it is crucial to address the challenges associated with biased data, transparency, and the potential exacerbation of existing issues within law enforcement. Implementing these technologies should be accompanied by strong oversight, safeguards, and efforts to promote fairness and protect individual liberties. By prioritizing ethical considerations and continually refining AI algorithms, we can harness the full potential of these technologies while upholding principles of justice, equality, and accountability in our pursuit of safer communities. As depicted in the image, the right balance between data privacy, and public safety should be drawn.



Towards Artificial General Intelligence

Recent Developments And Potential Risks

Utkarsh Sinha

In the history of evolution of life on earth, the emergence of humans represents a fleeting moment in a vast expanse of time. Yet, in the short amount of time we have been on the planet, we have managed to become the dominant species. The primary reason for this rapid ascent of the food chain is our intelligence. Given this fact, it is not surprising that we have always been fascinated by how intelligence works, and if it is something we can create or enhance. There have been several examples of fictitious creatures representing "artificial intelligence" throughout our history - from the myth of the Homunculus residing inside our body to modern-day stories like The Terminator. In the quest to develop artificial intelligence, the esteemed Professor John McCarthy initiated the modern version of this field as a research project at Stanford University in 1956. He believed that human intelligence isn't something that humans can never understand.

"I don't see that human intelligence is something that humans can never understand."
- Prof John McCarthy

The scientific community has worked relentlessly in the decades since then, and we have made major strides in the development of artificial intelligence. Today, we might be on the cusp of achieving the holy grail of AI – developing an "Artificial General Intelligence" (AGI). A machine which can perform any cognitive task we can conceive of - at or above human level. While current AI applications excel at specific functions, they lack general intelligence and cannot adapt to different tasks. For example, AI systems may outperform humans in one task but be outperformed by a child in another. Unlike AI, humans possess general intelligence, allowing them to perform a wide range of tasks and learn with fewer experiences. AGI aims to replicate this capability, enabling machines to learn and apply knowledge across various domains.

There is no consensus on when AGI will be realized. Estimates range from near-future to 100 years later to never, depending on the source. In any case, the development of such a machine will have profound impacts on our society – a society where cognitive and physical labour is abundant. It could create a self-improving machine that might lead to an intelligence explosion. The resulting artificial superintelligence (ASI) could either spell doom for our society or create a post-scarcity civilisation that would make present-day look like the Middle Ages in comparison. Borrowing the words of author James Barrat, the first true AGI may very well be "our final invention".

We will look at what form such an AGI might take and when based on the developments happening in the field at present. Would it be a completely digital entity, possess a physical body or interface with the human brain and enhance our intelligence? Or would it require some entirely new kind of approach that is not yet mainstream? How dangerous could it be in the hands of rogue actors, and what can governments do to mitigate this risk? Such questions, that existed only in the realm of science fiction just a few years ago, are becoming fundamentally important now. Let's try and explore these questions, starting with a look at the state of the art in the development of AGI.

"Of course, machines can't think as people do. A machine is different from a person, hence they think differently. The interesting question is, just because something is thinking differently from you, does that mean it's not thinking?"

– The Imitation Game

An Intelligent Digital Assistant

Imagine a future with computers that can take commands in conversational language and use the tools we use today to accomplish the task described to them. Indecisive about dinner? Snap a pic of your fridge for AI-generated recipes. Seeking a used cot on OLX? Your AI assistant will shortlist options, contacts owners, and schedules meetings. Office tasks like updating Salesforce or handling code documentation will become a breeze with one command to your AI office assistant. Imagine a visually impaired person walking down the aisle of a supermarket. The AI assistant could be their eyes, taking the video stream of the aisle on their mobile and directing them to the correct section of the mart to get what they need.

Microsoft launched Windows co-pilot: A natural language interface for all office tools

These capabilities may not be constrained to the realm of imagination much longer. Companies like Adept AI are already working on models that can use tools like Photoshop, Excel, Tableau etc. and work with you through a natural language interface. Microsoft recently launched Windows co-pilot, which aims to provide a natural language interface to all Microsoft Office tools. "BeMyEyes", a mobile app that connects visually impaired users to volunteers, is testing a beta version of its virtual assistant. Such applications show us the enormous positive impact that AGI can have on society - from freeing us up from the drudgery of work to helping members of society regain a sense of independence

It is important to note that these AI assistants show the capability of "multi-modality". They can see and identify objects in the real world, converse with humans in natural language, and understand and generate sound. The recent merging of the three modes of interaction is a big step towards achieving AGI. This, however, leads us to an open question in AI – can a true AGI be created without it ever interacting with the physical world? Some scientists believe so, and they are working towards what is known as "AI Embodiment".

AI Embodiment

Embodiment is the idea that an intelligent agent has a physical body that can interact with its environment. There is no consensus among researchers on whether embodiment is indispensable for AGI. Embodiment enables the AI to learn from sensory inputs and allows it to explore and experiment with physical objects in the real world. In a sense, we already have a limited version of embodied AI with self-driving cars. However, an embodied AGI must have greater capabilities than navigating roads. It should be able to interact with humans in their natural language and be able to perform tasks it was not specifically trained for. The embodied AI doesn't need to have a humanoid form either. Think R2-D2 and BB-8 from Star Wars or the robot assistants from Interstellar. They could be low-powered robots with designs suited for specific tasks.

Embodied AI could be of great use to humans in space exploration

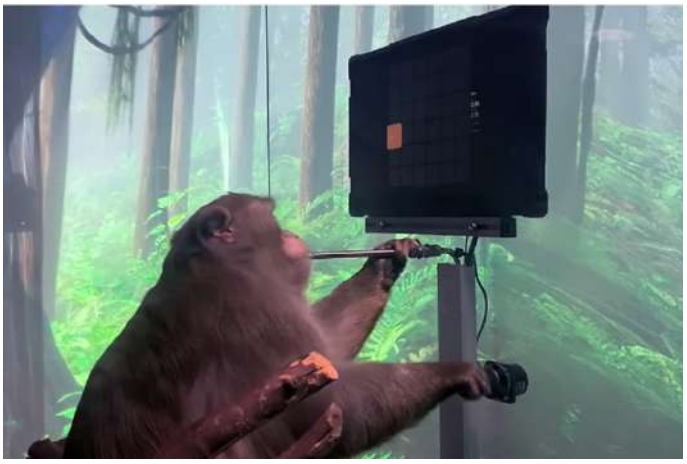
The humanoid robotics market is projected to reach $17.3 billion by 2027, driving intense competition for developing advanced humanoid robots. Google introduced the Palm-E (E for Embodied) model in March 2023, capable of solving diverse tasks across different robots and modalities. Around the same time, Open AI invested $23.5 million in 1X, focusing on androids collaborat ing with humans. Elon Musk announced Tesla Optimus at Tesla's inaugural AI day in August 2021.

Ultimately, the likelihood of AGI being realized as a humanoid robot appears high, given the increasing investments by tech giants in embodied AI systems. Whether it will be the final form remains uncertain and can only be determined with time. There is, however, a third possibility – that of humans merging with AI through a "brain-computer interface".

Brain-Computer Interface

A brain-computer interface (BCI) is a direct communication pathway between the brain's electrical activity and an external device, most commonly a computer or robotic limb. A BCI can translate a person's brain activity into external responses or directives, such as controlling a prosthetic limb with their thoughts. BCIs can also be used for augmenting or repairing human cognitive functions.

There have been successful experiments on Parkinson's patients that regained mobility through stimulation from a brain implant. Experiments on rats have shown the potential of implants being able to augment memory and transfer information using a brain-to-brain interface. Companies like Neuralink and Synchron are planning to start human trials of their implants.



Monkey playing video games with its mind using Neuralink's brain computer interface

However, these current applications only scratch the surface of what is possible. In the future, brain-computer interfaces could allow humans to leverage the strengths of digital computing, surpassing the capabilities of the unaugmented human brain. Today, smartphones act as extensions of ourselves, but limited data transmission between brains and mobiles hinders their full potential. Overcoming this limitation could enable direct access to machine intelligence, resulting in a collaborative "superintelligence" surpassing any previous cognitive abilities.

Would the creation of such a superintelligence be possible? It is too early to tell. The technology needed for humans to merge with AI might still be decades away. As we approach the development of AGI, the "misalignment problem" arises: its goals may not align with humanity's, posing an existential threat. BCIs can serve as a hedge, potentially granting us more control over AI by merging with it.

The approaches discussed till now assume the development of AGI in the form of a self-learning software. However, some scientists are working towards emulating the human brain through specialized hardware, in the nascent but promising new field of "Neuromorphic Computing".

Timeline and Form of AGI Arrival

The timeline for the arrival of the first true AGI is difficult to predict, with estimates ranging from as early as 2033 to as late as 2200, but most experts agree it will eventually happen. We saw last year that an increase in model size led to a breakthrough in performance of GPT 3.5 compared to GPT 3. Another 2 or 3 breakthroughs of similar magnitude may well lead us to achieving AGI in the form of a digital assistant, but the timing of such breakthroughs remain uncertain. Also, the ability of these chat bots to learn entirely new information by looking at new text hasn't been clearly demonstrated yet. Some argue that these

"One of the skills of AGI is not any particular milestone, but the meta skill of learning to figure things out. It can go and decide to get good at whatever you need."

– Sam Altman

bots are merely stochastic parrots imitating the patterns and language of the data they were trained on.

Embodied AGI is comparatively more difficult to build. The humanoid robots of today have limited generalizability in terms of the tasks they can perform. The potential job displacement in blue-collar work due to embodied AGI may also encourage governments to deter companies from making such an AGI on a mass scale. Hence, the development of AGI in this form is likely to take longer than in its digital form. Brain-computer interfaces have potential to become AGI, but their immediate use case is focused on curing neurological disorders rather than enhancing healthy individuals. The slow adoption and potential health risks may make this technology the least likely scenario for achieving AGI, although this approach gives us the best chance of achieving AI alignment. Other novel approaches, like emulating the hardware of the whole brain (neuromorphic AI), are still in their early stages and have a longer way to go to reach AGI compared to the previously discussed approaches.

Moving Forward: AGI Societal Impact and Risk Mitigation

Recently, there has been an active discourse among researchers and tech CEOs regarding the societal impact of AGI and ways to mitigate its risks. Proponents of AGI like Andrew Ng and Yan LeCun highlighted its positive contribution. On the other hand, notable figures like Elon Musk and Steve Wozniak have warned about its risks, calling for a pause on "Giant AI experiments". Geoffrey Hinton expressed concerns about the dangers of AI and subsequently left Google. Open AI and DeepMind published articles addressing existential risks and the need for governance and alignment of superintelligence. Bill Gates also posted on his blog comparing the development of AI to the development of the Graphic User Interface in its importance.

The pace of AGI development is accelerating, with advancements becoming more and more frequent. This is giving rise to risks like copyright infringement, learning from biased data, and helping rogue actors in cyberattacks and weapon development. These risks demand serious discussion and proactive mitigation. As we strive for AGI, important questions arise: Can big tech companies prioritize ethics and safety? Will the goals of a superintelligent being align with humanity's? How can AGI development be effectively regulated?

Big tech's proactive engagement with the government is promising, but uncertainty persists about regulations keeping up with AI advancements. Historical cases, such as banning of CFCs and nuclear power regulation after accidents, show that governments are not proactive in technology regulation. Artificial General Intelligence may not allow for such delayed regulation.

"Digital Superintelligence may be the biggest existential threat humanity has ever faced"

- Elon Musk

Open AI Charter: "OpenAI's mission is to ensure that artificial general intelligence (AGI), by which we mean highly autonomous systems that outperform humans at most economically valuable work benefits all of humanity."

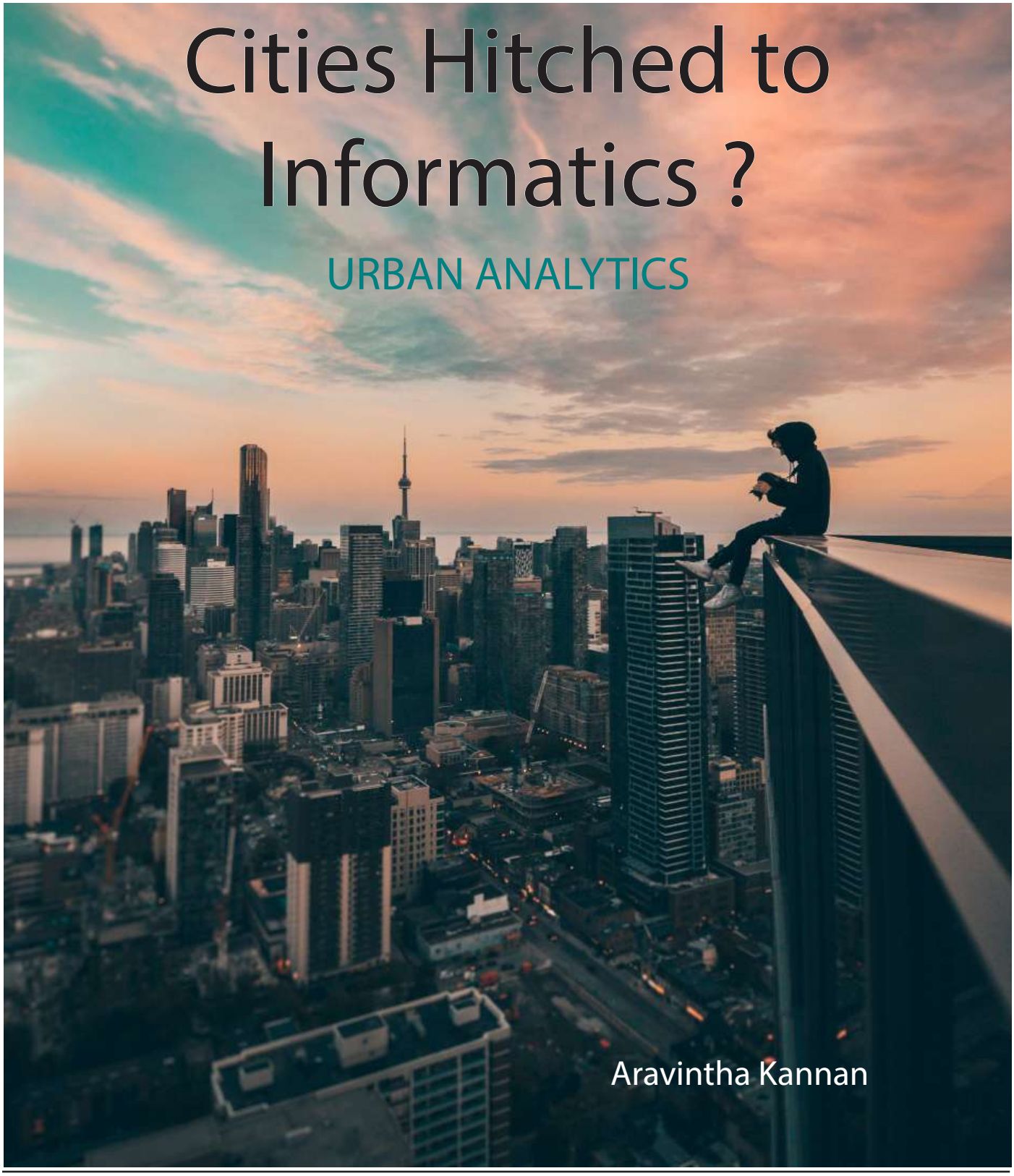
The development of AGI can be compared to the development of nuclear weapons. The country that develops it first will be at a decisive advantage. However, other countries will soon follow suit. The possible military capabilities developed as a result of AGI may be comparable to other weapons of mass destruction in its power. This warrants an agency like the IAEA to regulate AI at an international level. In fact, Open AI CEO Sam Altman has made such a suggestion himself.

Ultimately, like the Industrial Revolution granted us physical power beyond what the human body was capable of, AGI may grant us mental abilities beyond that which our minds are capable of. The industrial revolution was responsible for a quantum leap in the growth rate of the world economy. A similar jump in the growth rate of our economy may lead to a level of prosperity beyond anything we can imagine. Simultaneously, it may also lead to an increase in our destructive capabilities. Hence, we may be standing at an inflection point in human history. Our actions at this time might be remembered for generations in the future if we get it right. If we get it wrong, however, there might not be any future generations left to remember us.

The Prime Minister Rishi Sunak meets with Demis Hassabis, CEO DeepMind, Dario Amodei, CEO Anthropic, and Sam Altman, CEO OpenAI, in 10 Downing Street

Cities Hitched to Informatics ?

URBAN ANALYTICS



Aravintha Kannan

The continuous hustle for a better quality of life and upward social mobility over the past decade has accelerated urbanization in an unprecedented manner. Globally, this urban populace is expected to shoot up to 70% from 56% of the current figure in less than 25 years from now. In the present scenario with constrained geography, resources & infrastructure, delivery of public services to citizens in cities are lagging in technological solutions. What is the next paradigm to handle this? Is it our latest poster boy- AI?

The city never sleeps. It's a living, breathing organism with its own unique rhythms and flows. And yet, for many of us, the city can be a confusing place. We are constantly battling burgeoning traffic, struggling to find affordable housing, and feeling disconnected from our neighbours. But what if there was a way to make sense of it all? What if we could use data and technology to understand the complex systems that make up our cities? Hola, welcome to the world of urban analytics!!

Urban Analytics starts with location thinking- a superpower in understanding how objects relate to each other. Alongside, it uses urban data, which results from modelling the spatial effects of our existing systems in place. Having its primitive origins in mapping population densities today the field has leapfrogged miles and bounds to synthetic generation of data from models. Combining conventional domain knowledge and data, professionals adept at leveraging AI along with policymakers are able to solve major pressing issues for city dwellers. In its growing phase with huge potential, how broad is the relevance?

Man was known to be a social animal ever since stone age

Community

Man was known to be a social animal ever since the stone age, living in groups which today has transcended into living in dense urban spaces with solitary nature. Most of them hardly know their neighbours, but are well connected sitting in one part of the city and interacting online with their virtual friends in other parts of the city. As a consequence, a gold mine of social media network data is lying out there. Just a massive database isn't enough, the hierarchical metamorphosis of data to information to knowledge to wisdom yields real impact. This social media data is leveraged to gauge public mood using "Sentiment Analysis" and to extract topics/ themes of discussion using "Topic Modelling" which eventually helps in condensing large volumes of text into concise summaries. Based on this kind of real time monitoring of feedback from people, Governments can fast-track informed decision making to address bottlenecks for urbanites. Community structures characterise how the entities are arranged relatively with respect to each other, embedded within a spatial system. Detection of communities through graph-based clustering can aid in identifying the spread of epidemic diseases and allocation of necessary healthcare resources to mitigate it.

Inclusive AI

As a part of the developmental process, it's a common sight to observe "gentrification" which refers to the influx of wealthier residents and businesses into a previously lower-income/ working-class area. Though it has benefits of economic growth, it has a certain negative impact in terms of displacement of lower-income residents, loss of affordable housing along with widening socio-economic segregation. To mitigate this problem, a current solution in the pipeline is "Income Modelling"-analysis and prediction of income patterns and changes within a specific neighbourhood or urban area over time. Econometric tools such as multivariate regression with household demographics, local economic

Seamless integration of analytics and human ingenuity for journey planning will be the way forward

factors, Gini coefficient are used to estimate income. Such models can help in the early identification of areas in need of revitalization and provide targeted incentives in the form of prioritised investment to bridge the social cohesion.

Adaptive Mobility

In the heart of a bustling metropolis, navigating the labyrinth is full of CHAOS (Error 404: Polynomial order not found !!) for commuters. Unlike the random number generator, Urban working professionals have a constant route sailing to and fro their offices but do so in large quantums. Cities if not for proper planning would be crumbling to handle this large flow of people. Government and Policy Institutes in tandem with academicians are collaborating to capture both static and real-time data from public transit systems and utilise them for efficient design of services. To cite an example, Bengaluru's Municipal Corporation along with IISc have launched an app- "Namma BMTC" powered by AI Algorithms, where a lot of legacy data is used to predict the ETA of buses, which helps public transport users plan their journey in advance.

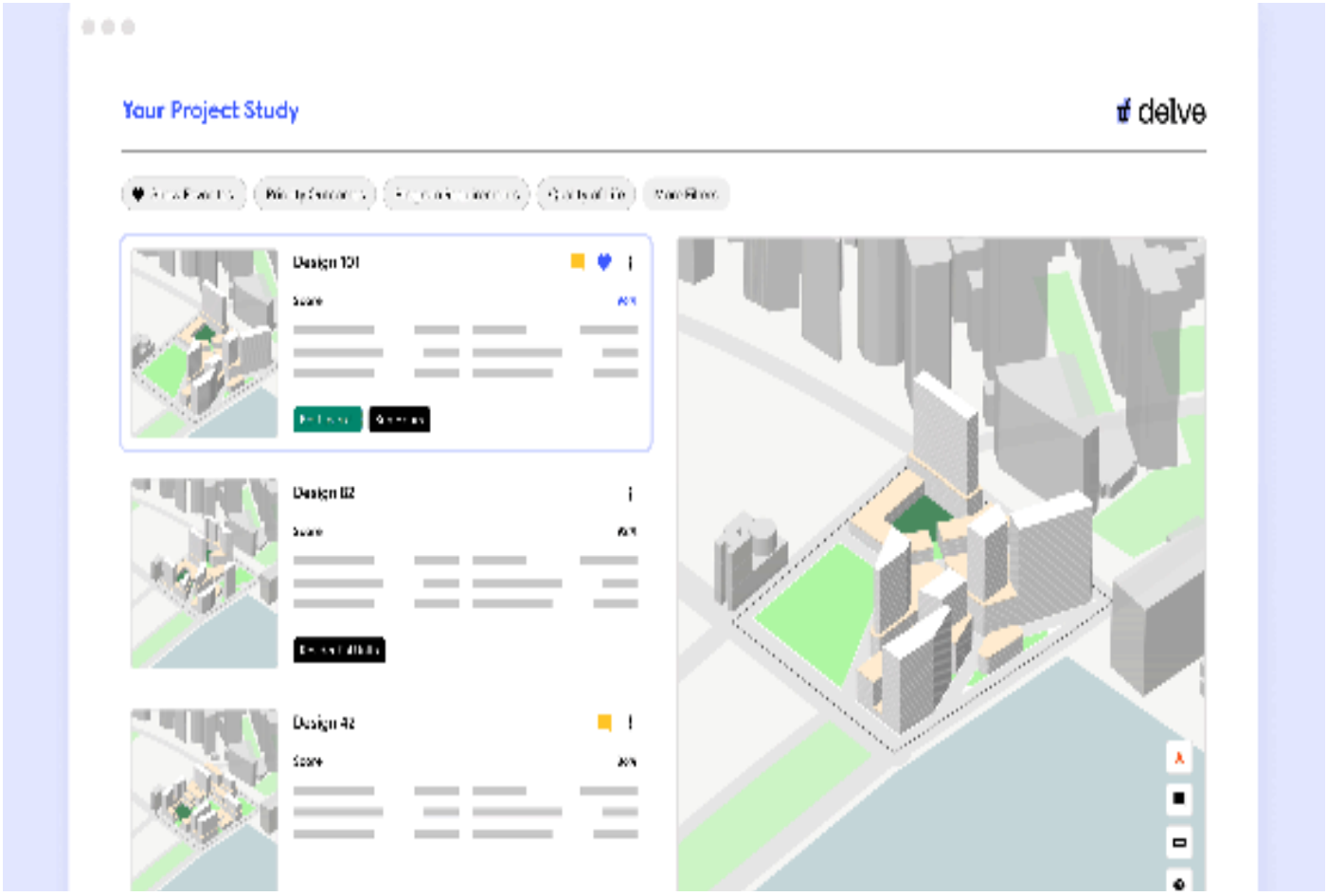
Currently, ML/ DL methods are being used to capture the spatial correlations and temporal dependencies involving origin-destination pairs, routing information and understanding mobility patterns. With the spurt of IOT devices, a lot of spatial trajectory data on moving objects- pedestrians, vehicles are collected. "Next Location prediction" is an upcoming area of research which involves forecasting the next location of an individual vehicle based on aggregate level with cross-sectional traffic volume and disaggregate level user centric speed profile. With responsive action to the results, it will help in minimising traffic congestion.

As this is a sequence problem, models such as Recurrent Neural Networks are used to predict the next location of the vehicle analogous to text generation. Notching it up one level advanced is "Attention-based Transformer", which concentrates on a certain part of network traffic state for the next cell generation. With inefficiencies in capturing data, techniques such as synthetic trajectory generation realistically reproduce mobility patterns. Closely associated to this realm is Generative Adversarial Imitation Learning, which is used for modelling pedestrian behaviour. Trajectory data collected through motion capture systems are encoded as continuous sequences of positions/velocities and trained using a policy generator. A discriminator network is trained to distinguish between expert trajectories and generated trajectories. Broadly, seamless integration of analytics and human ingenuity for multi-modal journey planning will be the way forward from the present. Having discussed the impact of population on mobility, will existing cities suffice for the future?

Generative Planning

Satellite cities with their own employment bases have been emerging as potential solutions to tackle the problem of rising population in cities. The discipline of urban planning comes into picture here- creating urban spaces incorporating buildings, environment, traffic. As Generative AI is the bandwagon that every industry is looking to hop on, then why should urban planning be left alone? While designing a new neighbourhood, planners, architects and developers weigh innumerable competing factors such as shadows, green cover, accessibility index. Combining machine learning and computational design, it has now become possible to generate different feasible comprehensive scenarios at scale, by taking the different objectives into consideration and assessing the necessary trade-offs. Sidewalk Labs, an innovative company backed by Google is working on a whole lot of interesting set of projects in this field. One of their main products- Delve, provides different iterative options as suitable designs for a particular set of constraints as input. The laborious process of going through thousands of parameters and manually creating designs is a cakewalk for AI. Apart from faster calculations, it helps in increased utilisation of residential areas and open spaces. But with all this technological progress, is the city agnostic or empathetic towards the environment?

Sidewalk Labs, backed by Google, is working on problems involving Generative Planning



Delve provides varoious iterative options as suitable city designs for a particular set of constraints as input

Green Intelligence

Eco-friendliness will become a core tenet of future cities, where energy-efficient office buildings will adjust the climate accordingly and lightings will tune in based on sitting preferences. With urbanisation, arises the commitment to manage the generation and distribution of energy in a sustainable manner to reduce the carbon footprint. Green roofscapes with photovoltaic panels are seen as an emerging trend in this aspect. They along with fulfilling the climate change pledges, provide insulation to mitigate what is known as the Urban Island heat effect which is a phenomenon of cities replacing natural land cover with dense concentrations of pavement, buildings, and other surfaces that absorb and retain heat. This effect increases energy costs, heat-related illness and mortality. Green roofs provide thermal comfort to urban residents and reduce excessive air-conditioning energy. Having stated their utility, what if their spatial distribution could be studied to evaluate the carbon offset capacities of cities? Fine-tuning computer vision models such as Convolutional Neural networks (CNN) performing image boundary identification, instance segmentation for roof topology detection is a solution proposed by the Urban Analytics Lab at NUS Singapore. This helps in identifying optimal locations for solar panel installations, estimating energy generation capacity, and promoting the adoption of renewable energy sources. And is it fine if the environmental health alone is fortified, what about the breathing apparatus of humans living there?

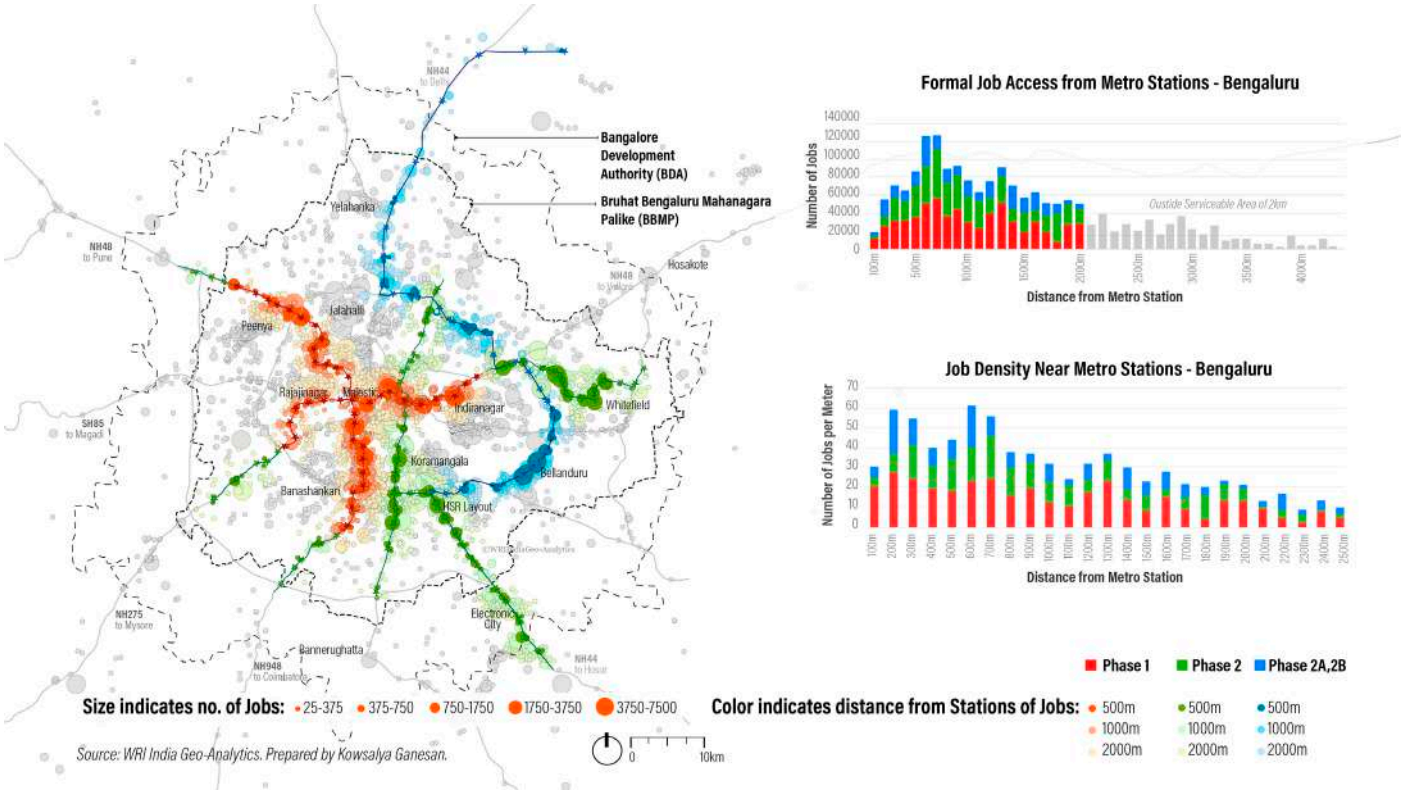
There's this popular joke circulating on the internet- "Mr. India doesn't need his watch in Delhi. No one can see anything anyway". To combat air pollution, the adoption of streaming analytics data from gas sensors are used to monitor air quality index. With EVs looking to gain traction over the next few years, Governments will have to spend extensively on building the charging infrastructure for their mainstream adoption. Models to estimate charging demand using various features such as Population Density, EV adoption rate, charging duration, real time shall become crucial to identify the ideal locations to install these charging stations. Conducting Impact analysis through carbon accounting data can assess the monetary value of social benefits in carbon reduction.

Road Ahead

With urban data being a boon in many ways, it has its own pitfalls as well. First, it undermines the privacy of citizens, subjecting them to extended personal surveillance. This data if hacked by nefarious people will lead to crimes of all sorts. Emerging new-age quantum-based cryptographic methods are trying to address this issue. Another area of ethical concern is algorithmic biases in models used for urban data. Then, conducting regular audits to monitor the model performance and collecting "representative data" is necessary for ensuring fairness. As urbanisation evolves, the amount of data generated will increase exponentially. Latency in terms of data transmission and scalability in terms of infrastructure will pose a challenge. But with advancement of cloud technologies, we can dynamically provision the computing power, storage and networking resources to handle large-scale data processing tasks. Albeit, finally one has to be careful with deploying machine intelligence at urban spaces of intersections with humans, where safety of these systems is still a grey area. Recently in the city of Pune, an artificial neural network pipe burst like a dormant volcano. However, robust testing protocols are being developed to ensure physical safety of such systems. In essence, with responsible governance mechanisms, urban analytics will play a positive role in reimagining our future cities as thriving hubs of innovation and social well-being.

Bengaluru: Namma Metro engages World Resources Institute as technical advisors in two projects

According to BMRCL, WRI India is a global research organisation focusing on building sustainable, liveable cities and working towards a low-carbon economy.



Corporations, along with IISc, have launched an app- "Namma BMTC" to predict ETA of buses

JOSHUA STARMER
Founder and CEO, StatQuest

Dr. Starmer, a visionary data scientist and educator, holds a strong background in computer science and music theory from Oberlin College, as well as a Ph.D. in bioinformatics from North Carolina State University. He has also worked as a postdoctoral fellow and assistant professor at the University of North Carolina at Chapel Hill. His commitment to knowledge sharing is evident through his creation of StatQuest, an educational platform with over 936K subscribers.

AINA: So Dr. Starmer, first we'd like to begin with your journey. We are intrigued by your diverse background as a founder, educator, and YouTube creator as well-who's focused on statistics and machine learning. We are curious to know what motivated your academic transitions, from undergrad in music and computer science to pursuing a Ph.D. in bioinformatics. Could you share your experience regarding these academic transitions and how they have shaped your journey thus far?

I grew up playing music my whole life. I grew up playing the cello, which is, if you don't know, it's a large instrument related to the violin. And I just loved music, and I wanted to be a musician. I thought that would be a great career option for me just because I was so passionate about it. But I went to college, and every summer I would apply for jobs, you know, to make some money. And every summer I was offered an unpaid position with music or even a position where they're like, you can pay us. And I'm like, are you kidding me? Or I was offered like a paying computing job, like a programming job. And to be honest, programming - I mean, I love music and I'm very passionate about music - but I also love programming. Programming is like, it's super fun. It's like solv ing little puzzles. I love this. I love that feeling of working on a programming project and, you know, getting it done. And so to me, it really was sort of like a little bit of a coin toss. I maintained both, and I still play music and was able to get a degree in music and a degree in computer science. So, it was not a choice. I had to do both.

AINA: Recently, we came across Google developing a music language model that could generate music that remains consistent over several minutes as well. How do you think this changes the scene for musicians like you?

That's interesting. I could talk all day about these things. At various points in my career, I've actually been a professional musician and a professional composer. I've written music for television commercials. I've written, performed and composed music for movies, for dance companies. I've done a lot of that. And what I discovered while I was doing that was, a lot of the time people don't want new creative music. As a composer, I found this very frustrating. What they really wanted was their movie to sound just like some other movie, or they wanted their dance music to sound just like somebody else's dance music. And it was one of the reasons why I got out of that. I found it very recreational and sort of imitative rather than creative. And if imitative stuff is what they want, they could spare a lot of musicians, a lot of agony by just creating it. I know I'm probably making light of it. There are a lot of jobs that are on the line here. It's a much more complicated thing than the creative people versus the imitative people because there are lots of people that just like to make music.

I'm an optimist, and so I'm going to look at this thing from well, what good can come out of this? And it could be that - as a result of it being so easy to create imitative content, people will now put a premium on creative content instead of just trying to make everything sound the same or sound like whatever the next hit or last hit was.

"The positive outcome (of MusicLM) could be that - as a result of it being so easy to create imitative content, people will now put a premium on creative content instead of just trying to make everything sound the same or sound like whatever the next hit or last hit was."

AINA: Any reason for using those opening songs in your Statquest videos?

I've found that when people hear my little silly song that's unprofessional, and clunky which doesn't even sound great, it has a calming effect. At least for some people. They're stressed out. They've got a big test ahead of them. They've got a job interview. They've got a lot of things that they're worried about and they're watching my video to practice, study, or get prepared for something they're really stressed out about. And my silly songs sort of help them relax. That is why when I start the video I start with a jingle and BAM, they feel relaxed.

AINA: Dr. Stammer, you're a Ph.D. in bioinformatics. We have had basic neural networks to language models with memory retrieval now. What kind of further developments do you expect to see in the future of AI being inspired by biology?

As far as I can tell is for AI to actually get further and further and further away from biology. Take neural networks. The original concept was like- Oh, we'll make something sort of like a neuron. And they did. They tried to model a basic neuron, using something called a sigmoid curve in those equations of how a neural network should work.

"[Advanced deep learning architectures] are gradually shifting away from biology"

Well, the reason why they incorporated that was the sigmoid did a relatively good job with that S shape. Because that's sort of the way neurons behave when you trigger a neuron. That's great. But it turned out that that S-shape was getting in the way of neural networks doing what they needed to do. When you use that S-shape, neural networks could not go could not become what's called deep, in that they were limited in terms of the kind of problems they could solve with the S-shape. And so, about a little over 20 years ago, they said - OK, , let's get rid of the S-shape. And they switched to .something called a ReLU activation function, which is a bent shape. It doesn't saturate and it's very non-biological in how it works.

Once they got rid of that, all of a sudden, neural networks started performing much better. They could get much deeper and they could approximate much more complicated functions and much more complicated data sets. That was like a big step away from biology. Now we've got square root functions. We're taking dot products left and right. We are starting to add layers of functions that no longer have anything to do with biology and how the brain works. So we've taken a step away from biology. We created transformers, which have given birth to the LLMs that we have now. And so, from my perspective, what I've seen is a relatively gradual shift away from biology.

AINA: What was the inspiration behind your YouTube channel and how did you start?

I was at the University of North Carolina Carolina in Chapel Hill working in the genetics department, and it's an academic laboratory environment, which means students are coming and students are going. And when they come, they need to learn a lot of things, some of them more than others but most people in the lab were genetics people. They did not always have great statistical backgrounds. They had great genetic backgrounds, which wanted them. But at least in genetics, it's important to also have some statistical background. So, I was teaching people statistics.

My motivation for teaching my coworkers statistics was a couple of things. One is just like computing and music. I am really passionate about statistics. Statistics is sort of a magical field to me in that, when we grow up and we take normal math classes, everything is math. It is hard to describe. Math is almost like a pure thing that happens in the heavens, right? It's like, you could say, one plus one equals two and it always equals two. And that two is the same for all time and eternity and it never changes.

But in life and reality, one plus one does not equal two. Say as someone gave me two French fries. Would those two French fries weigh the same as two other French fries? Would they be the same size? Would they be just as salty and fresh?

Life is not the math we were taught. One plus one doesn't always equal the exact same two as we thought we were getting. Statistics is the only thing I know that actually tries to deal with that. And not only does it try to deal with it, it tries to give it, make a strength out of it. Statistics is somewhat magical that way. So, long story short, I wanted to share my passion for statistics just like as a musician, I wanted to share my passion for music by playing music for other people.

AINA: When you start making videos, do you follow any specific strategies to break down a certain topic? Say, you are doing an LSTM topic now, do you follow any strategy for that?

For any concept that I want to

teach - neural networks, basic statistics - you name it. The fact of the matter is there are already tons of resources out there that teach the exact same stuff. There are websites, there are blogs, videos, podcasts. Maybe you have a Hollywood feature film or a dance routine or something like that which tries to express these concepts.

Unfortunately, a lot of the educational material on the internet right now is imitative. I do not know how many videos about linear regression are there that talk about it in the exact same way. And the problem with this sort of imitative style of teaching is, say I watch a video and it does not make sense to me and I watch another video that is imitative, and it's basically the same video. That video is not going to make any sense to me either. Maybe they have got fancier graphics, but I'm still going to be confused, right?

And so my goal when creating material is not to be imitative. My goal is to be creative and to say the things that other people don't say. I don't really watch other people's videos before making mine because ,then I may end up imitating them.

AINA: Can you share any memorable feedback or success stories from your students or viewers who have benefited from StatQuest?

Back really really early on, someone from Indonesia watched a video and posted a comment saying that they just won a data science contest because of what they learned in my videos. And that was probably one of the most inspiring things that I ever read in terms of my work. It was transformative and it was the thing that opened my eyes to the fact that it wasn't just my co-workers watching what I was doing. All of a sudden, what I was doing wasn't just meaningful in a really limited context. It was meaningful in a global context when someone on the other side of the world just won a data science contest. I was like if that doesn't inspire you to want to make more and be better at your job, I don't know what will. Once I realized that the impact was much greater, it became something more than 'Oh, I'll just do this in my spare time'. But then I was like, wait a minute, no, this is important. This is something I need to be doing because it's bigger than just this group. That is when I started focusing more of my time and eventually, I had to leave my job so I could spend as much time as possible making these videos. But that was really the first one that set everything in motion.

"Math is like a pure thing that happens in the heavens.... But in life and reality, one plus one doesn't equal two."

I want everyone to win a contest or get a job or graduate or get a good grade. I want those things to happen to everybody. It's kind of like having a small role in somebody else's success. It's like I feel it too. I feel the joy and I feel the victory.

AINA: When you announced that your book is coming to India, it got many of us excited. How did you realize that the content you uploaded or your book's content was perfect for everyone to understand any concept quickly and adequately?

You know, I just wasn't very gifted. I was not a natural storyteller and the thought of writing a book just seemed impossible. How could anyone - or especially me - write 300 pages about something? Just writing, like, language ,was so challenging for me. But I realized that I actually didn't have to write a book. I could draw a book. I could make cartoons and I could have a Stat Squatch and I could have the normalosaurus and I could have all these little things that make it very graphical instead of just being a bunch of words. So,once I realized I could do that, I was like, oh, wait a minute. I can do this!

Writing that book was another pivotal point in my career, because I went from someone who refused to believe it was possible for me to write a book to actually writing a book and loving every minute of it. I mean, I love making videos. I love the stuff I make up for YouTube. It's super fun. But one thing I didn't realize until I wrote a book was that it gives you some room to step back just a little bit and see the big picture, and you can see how everything's related.

Currently, I am working on another book on neural networks and deep learning and its use cases like chat GPT and stuff like that. I have been researching it for the past six months and I'm almost done. I'll hopefully be done at the end of June and then, starting in July, I am going to spend the next six months just writing that book.

AINA: People from diverse backgrounds are aiming to ride the AI wave. How would you suggest people with little math background pick up concepts in machine learning which require understanding the maths behind them?

I would highly recommend my book: StatQuest Illustrated guide

to machine learning or watch my YouTube videos. I think they're great educational resources. But what are other ways? There's a lot. Just be persistent. The information is out there.

One thing I found, and I'm starting to recommend to people, is to start with anything that gets you some sort of intuition about what's going on. And then once you have that intuition, believe it or not, you can actually read the original manuscripts because a lot of this stuff is published and freely downloadable as PDFs. And I think that might be a really great way to learn, because you have the concepts in mind. If you have some intuition of what's going on, you have some sense of what the language is, what the terminology is that they're going to use.

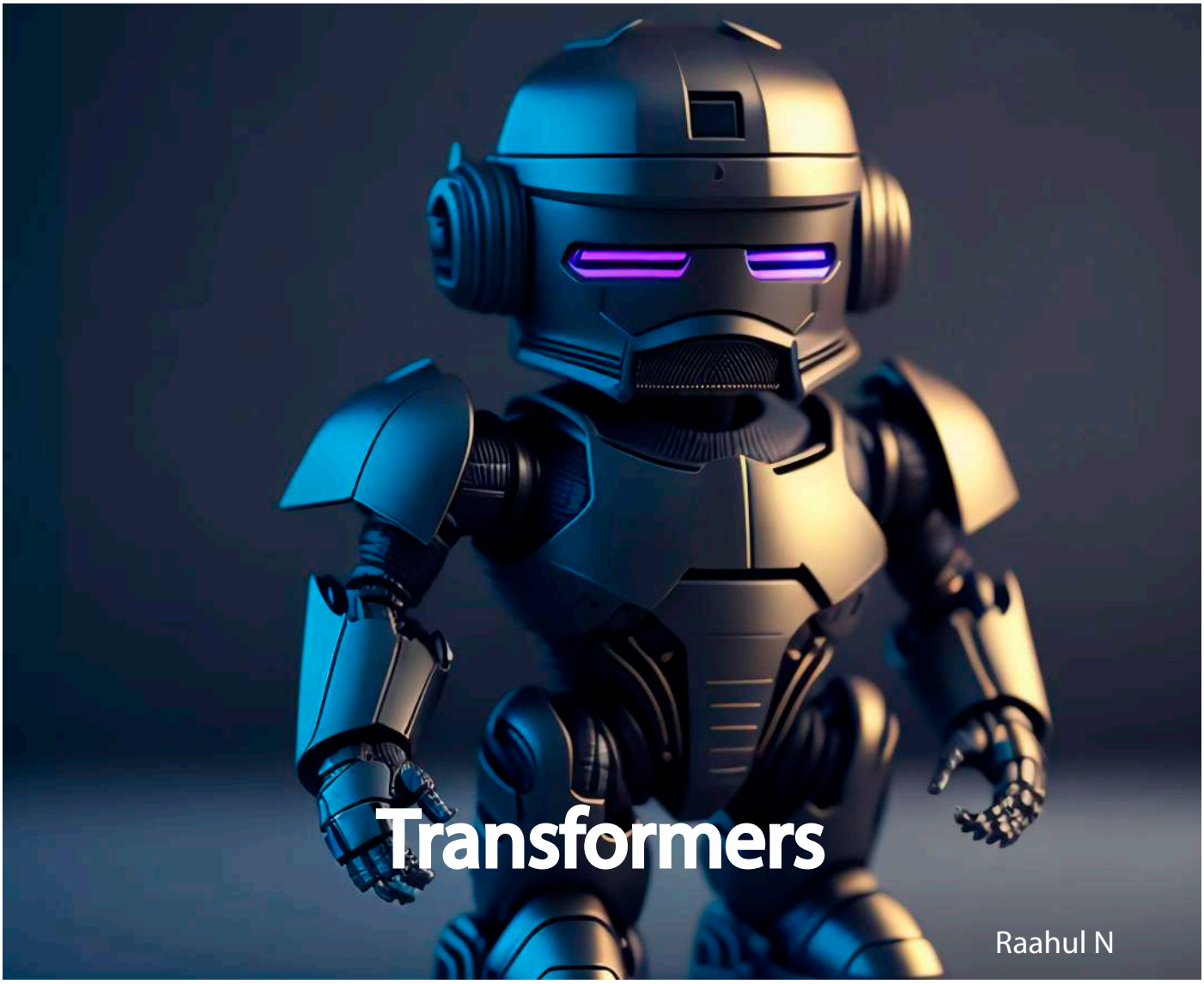
"All of a sudden what I was doing wasn't just meaningful in a really limited context. It was meaningful in a global context when someone on the other side of the world just won a data science contest. I was like, if that doesn't inspire you to want to make more and be better at your job, I don't know what will."

I know it's very buzzy and I know it's a lot of work, but AI is very buzzy and it's a lot of work for people to come from different backgrounds and get on it because there's a lot of catching up that needs to be done. That's my favorite approach. You can start with anything but start somewhere with the goal of trying to understand the original manuscript.

AINA: Can we expect your deep learning book in India anytime soon?

You could expect it basically a year from last month. I'm hoping to write it in six months and then it'll take probably another four months to edit. Editing is super important, because the first draft is rough. I have to read it a bunch of times. Make professional editors read through it as well. I have to do an early release to a select group of people who can read it and give me feedback. The editing ends up taking almost as long as writing the original manuscript, but I'm hoping that May 2024 should be that month.'

I am really excited about this book, because it will be different from my previous book. It is going to have a lot of stuff about neural networks, AI, deep learning, and large language models, but it will be having coding examples as well for everything. It won't just be dry theory. There will be QR codes that take you to a worked-out example of how it all goes. You could also plug in your own data and just play around with it. That's the plan. Sounds exciting doesn't it?



Attention Here

It all started with "Attention is all you need" in late 2017. Vaswani and a team of researchers from Google and the University of Toronto released a paper on this topic, introducing the transformer architecture. This architecture eschewed recurrence and relied entirely on an attention mechanism to transmit information across the input sequences and draw global dependencies between input & output. This turned out to be revolutionary just in a couple of years.

Transformers paved the way for parallelising the training process, meaning using more GPUs and less time to train. The Self-Attention mechanism that they introduced also captured contextual meaning of words better leading to superior results on NLP benchmarks. Needless to say, the world's attention was on Transformers.

It's fine tuning

Training a large language model is immensely expensive computationally and as a consequence economically. With the impressive performance of Transformers, big techs poured billions into this to have the best language model. Either as a race to capture the new search market or to revolutionise the AI assistance space, this paved the way for off-the-shelf models that can be used to suit our specific needs. Models based on transformers like BERT and T5 (both from Google) performed exceptionally well in numerous NLP tasks. Developers could just train/fine-tune (retrain the last few layers) BERT with their limited computation power and data to come up with their own model. Some notable big players are

- Facebook's LLaMA
- OpenAI's GPT
- Google's Bard

A Look at the frontier of exponential Progress

Generative Agents: Interactive Simulacra of Human Behaviour:

A recent revolutionary study by researchers from Stanford and Google brings forward the concept of generative agents which are essentially computational software agents. This study is considered to be a turning point in AI as the generative agents are capable of simulating 'believable human behaviour'.

What are generative agents? Based on this study, generative agents are computational software agents that use AI to simulate believable human behaviour. These agents draw on generative models to imitate individual and group behaviours. These behaviours are human-like and are based on their 'identities, environment, and changing experiences'. Just imagine the things we could do with a well-represented simulation of human behaviour.

Ever heard of AutoGPT?

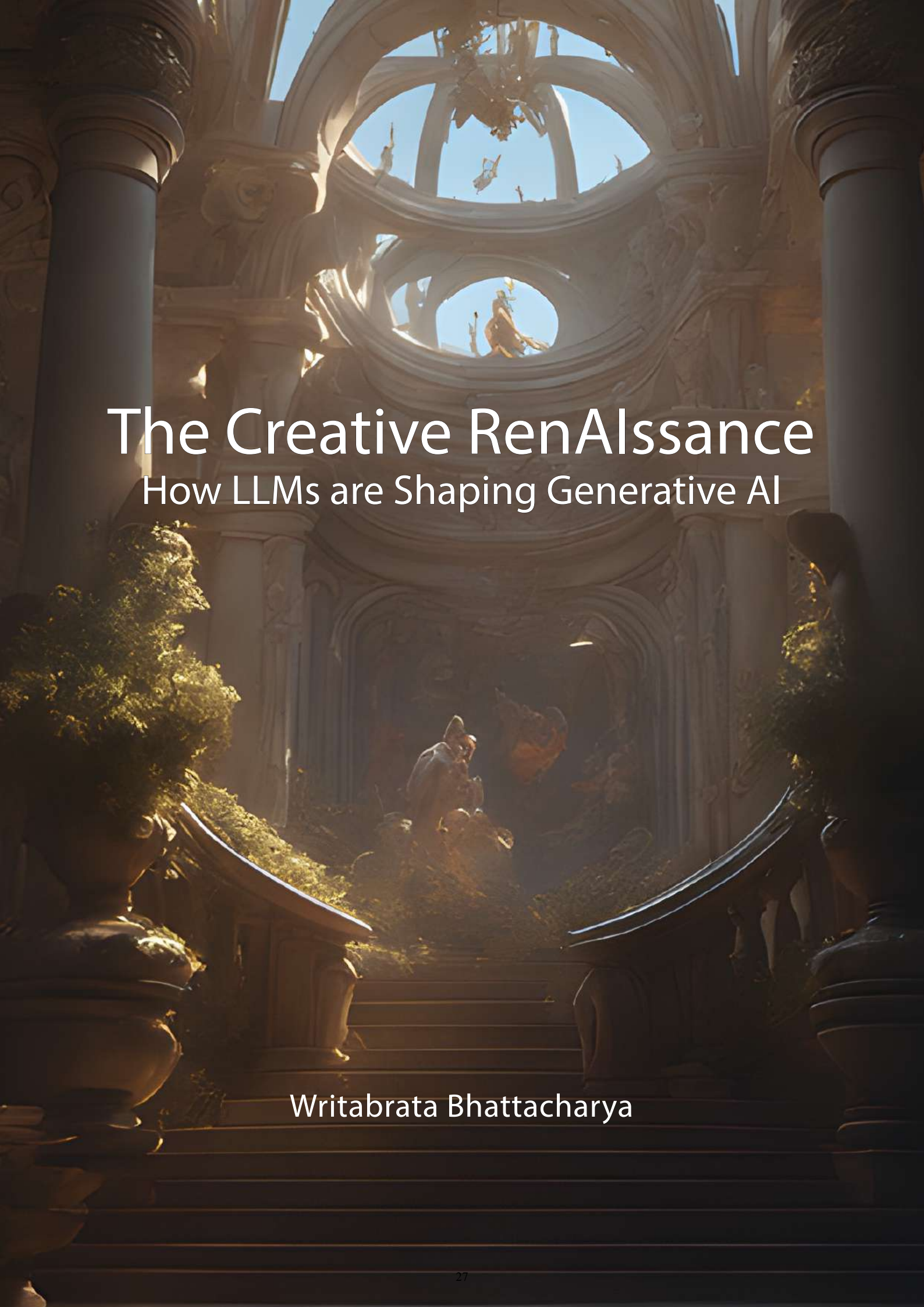
"Auto-GPT is an experimental open-source application showcasing the capabilities of the GPT-4 language model. This program, driven by GPT-4, chains together LLM "thoughts", to autonomously achieve whatever goal you set. As one of the first examples of GPT-4 running fully autonomously, Auto-GPT pushes the boundaries of what is possible with AI"

These are autonomous agents which utilise the language understanding of GPT to perform any assigned task. For example, I could run this on my laptop to create an agent who writes articles for a student-run AI magazine. Her goals could be to search the internet for the latest trends in AI and summarise them in a 1000-word article, and I would have an article in (maybe) minutes.

Integration of AI into any workspace/software

Chatbots (needless to mention), AI assistants, Search engines-in any software you can think of, the "help" option is going to be integrated to an autonomous agent which could understand & complete the requested task for you.

These could just be the tip of the iceberg. If this is the progress in about 6 years, the coming decade definitely has a lot in store for us. Although billions of dollars are already locked up in billions of parameters in a bunch of transformers, what can we say, a novel architecture could be disruptive enough to convince the big tech to pour billions more into a new architecture or not.



The Creative RenAIssance

How LLMs are Shaping Generative AI

Writabrata Bhattacharya

AI's Uncharted Voyage

Back in 1983, Claude Shannon, the renowned American mathematician and computer scientist, who is credited with inventing Information Theory and introducing Shannon's Entropy, expressed a fascinating vision for the future: Today, we stand on the verge of turning that vision into reality, witnessing a remarkable surge in the utilization of artificial intelligence over the past decade. From its origins as a purely academic pursuit, AI has rapidly evolved into a powerful force driving various industries and profoundly impacting the daily lives of millions of individuals.

"I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines"

- Claude Shannon

In the world of technology and artificial intelligence, we've recently achieved remarkable milestones, developing AI systems that can learn from vast amounts of data, ranging from thousands to millions of examples. These cutting-edge advancements have revolutionized our understanding of the world, unlocking creative solutions to intricate problems. Thanks to these large-scale models, we now witness AI-powered systems effortlessly grasping human speech and written language. Think of the natural-language processing and understanding programs that have become a part of our daily lives, like digital assistants and speech-to-text applications. Beyond that, these extensive datasets, containing everything from the masterpieces of renowned artists to the entire collection of existing chemistry textbooks, have opened new frontiers in generative models. These models are now capable of producing awe-inspiring artworks inspired by specific styles or proposing novel chemical compounds based on the history of scientific research. The possibilities seem limitless as AI continues to redefine what's achievable in the realm of entertainment and technology.

In the realm of AI, while numerous systems are making a real impact on practical issues, their development and implementation demand substantial time and resources. To address specific tasks effectively, having a comprehensive and well-labelled dataset is crucial. Otherwise, human effort is required to painstakingly locate and label suitable images, text, or graphs. The AI model must then learn from this dataset before it can be tailored to your unique needs, be it language comprehension or discovering new drug molecules. Additionally, training a large-scale natural-language processing model can have a significant environmental impact equivalent to running five cars throughout their lifespan.

Rise of Foundation Models

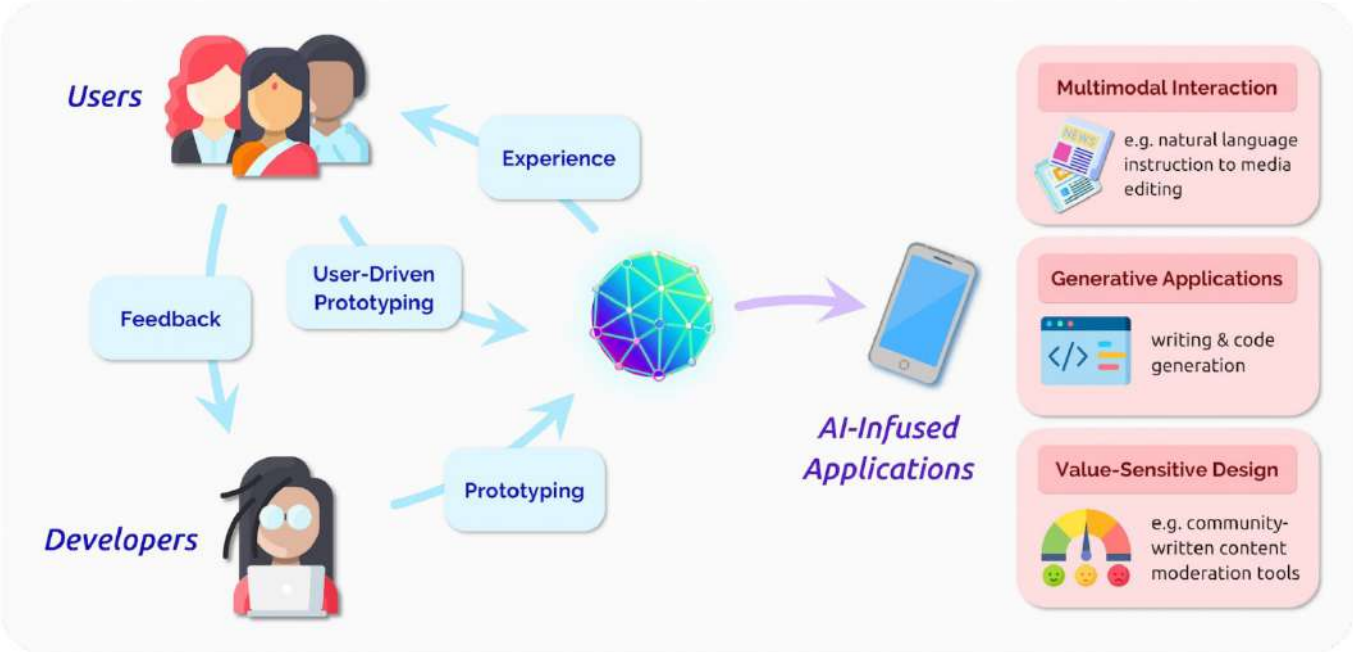
The next phase of AI is poised to revolutionize the dominance of task-specific models that have prevailed thus far. The future lies in the realm of foundation models, which are trained on extensive sets of unlabelled data, enabling them to perform diverse tasks with minimal fine-tuning. The concept of foundation models was popularized by the Stanford Institute for Human-Centered Artificial Intelligence through a comprehensive 214-page paper published in the summer of 2021.

Early glimpses of the potential of foundation models have been witnessed in domains like imagery and language, with pioneering examples such as GPT-3, BERT, and DALL-E 2 showcasing their impressive capabilities. These systems can generate complete essays or intricate images based on brief prompts, even though they were not explicitly trained for those exact tasks or image generation in that precise manner.

Through self-supervised learning and transfer learning techniques, the model can apply the knowledge gained from one situation to another, much like how learning to drive one type of car allows for an easy transition to driving other types of vehicles with minimal effort. This advancement significantly boosts the efficiency and versatility of AI systems, enabling them to tackle new challenges without extensive retraining or reliance on specialized models for each task. As a result, the horizon of possibilities for AI applications expands, pushing the boundaries of what can be achieved with this transformative technology.

"The question of whether a computer can think is no more interesting than the question of whether a submarine can swim"

- Edsger W. Dijkstra



Generative AI: Unleashing Creativity

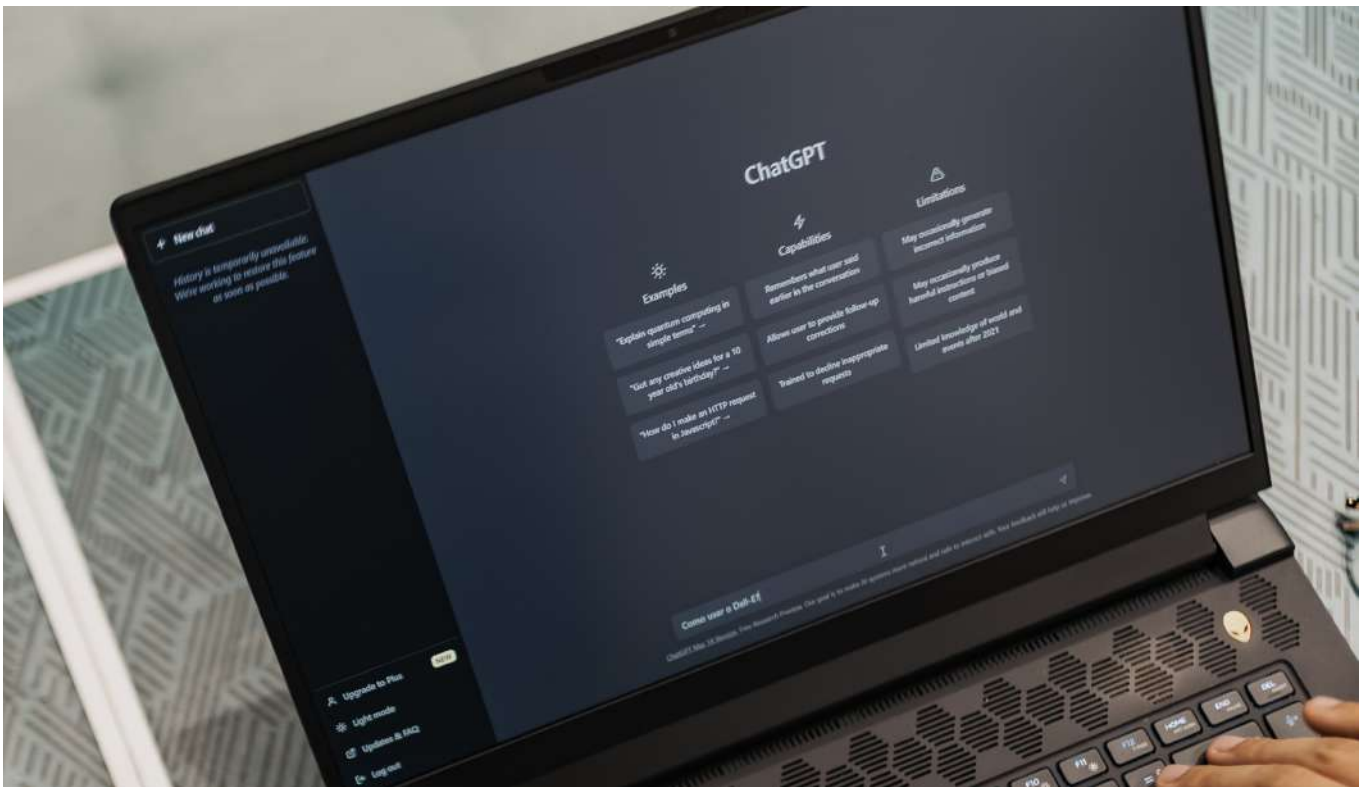
The emergence of foundation models has paved the way for a transformative branch of artificial intelligence known as Generative AI. This incredible technology enables AI systems to create new and original content in different forms, like text, images, audio, and even synthetic data. Just like you draw or write stories, generative AI can do that too, but it's incredibly fast and can produce a multitude of amazing creations in a short span of time. It's like having a super-smart creative assistant that helps us explore fresh ideas and bring forth fun and exciting possibilities in the world!

In contrast to other AI approaches that primarily focus on recognition, prediction, or classification tasks, Generative AI centres its attention on creating new data by leveraging patterns and structures learned from training data. In other words, it abstracts the underlying pattern related to input and uses that to generate similar content.

To illustrate this concept, consider a generative AI model trained on a dataset of human faces. This model has the capacity to produce lifelike and original faces that bear resemblance to the examples it was exposed to during training. Similarly, a generative AI model trained on a collection of poems can generate new poems characterized by a similar style or theme, showcasing its creative potential.

Generative AI holds significant implications across a multitude of fields, spanning art, entertainment, design, and scientific research. In the realm of video games, it can be employed to generate realistic graphics that enhance players' immersion and visual experience. Additionally, generative AI can generate synthetic data that aids in training other AI models, facilitating the development and advancement of AI technologies. Moreover, it can serve as a valuable tool in creative endeavours, assisting in writing or music composition tasks by generating novel ideas or compositions. In the realm of scientific research, generative AI can contribute to drug discovery efforts by generating new molecular structures with potential pharmaceutical applications.

In recent times, two notable breakthroughs in the field of AI have captured significant attention: Large Language Models (LLMs) such as ChatGPT, Bard, and PaLM, as well as Audio Generation Models (AudioLMs) like MusicLM.



Generative AI enables AI systems to create original content in different forms in a short span of time

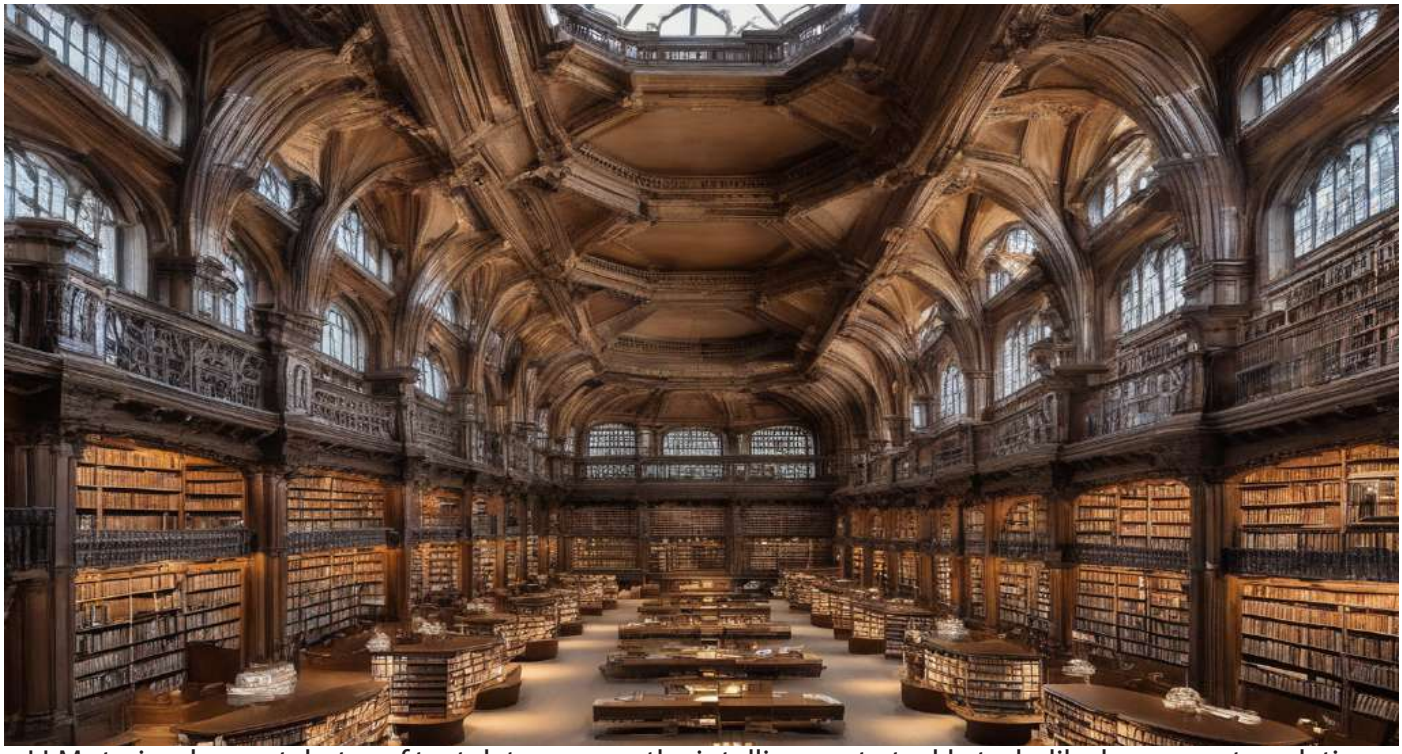
Large Language Model

A large language model is a trained deep-learning model that understands and generates text in a human-like fashion. It is like a very smart robot that knows a lot about words and can help us with reading and writing. It has learned from many books and stories, so it understands how sentences are made and can answer questions or even create new stories for us. It's like having a clever friend who can talk with us and help us learn new things about language and words.

These models can be understood by breaking down their key concepts into three components: "Large", "General purpose", and "Pretrained & Fine-tuned". "Large" refers to the vast size of the training dataset, sometimes reaching the scale of petabytes, and it also pertains to the parameter count, which represents the memories and knowledge the machine acquires during model training. Parameters determine the model's proficiency in solving specific problems, such as text prediction.

These models are called "General purpose" as they are versatile enough to address common problems. This concept arises from the universality of human language, which transcends specific tasks, combined with resource availability constraints. Only a select few organizations possess the capability to train such large language models with massive datasets and a tremendous number of parameters. "Pretrained and fine-tuned" refers to the process of initially pretraining an LLM on a vast general-purpose dataset and subsequently fine-tuning it for specific objectives using a much smaller dataset.

There are several benefits to utilizing large language models. Firstly, a single model can be applied to various tasks. These LLMs, trained on petabytes of data and boasting billions of parameters, possess the intelligence to tackle tasks like language translation, sentence completion, and text classification. Secondly, LLMs require minimal domain-specific training data when tailoring them to address specific problems. Even with limited training data, LLMs can perform reasonably well and exhibit the ability to recognize patterns that were not explicitly taught during training. Lastly, the performance of LLMs continues to improve as more data and parameters are added.



LLMs trained on petabytes of text data possess the intelligence to tackle tasks like language translation

Over the past few months, LLMs have left a strong impression in various fields. From crafting poetry to helping with vacation plans, we're witnessing impressive progress in AI's abilities, creating value for businesses and individuals alike. It's even quite possible that the very article you are reading right now was generated by AI itself.

Pathway Language Model

One prominent example of a Large Language Model is PaLM, developed by Google. PaLM, short for Pathways Language Model, is a dense decoder-only transformer model that leverages the innovative pathway system, enabling Google to train a single model across multiple TPUs. This model exhibits proficiency in various domains, including mathematics, coding, advanced reasoning, and multilingual tasks such as translation.

During its training, PaLM was exposed to a diverse range of data, encompassing scientific and mathematical information, 100 spoken languages, and over 20 programming languages. It serves as the underlying technology for Google's workspace products, Bard, and the PaLM API. It is possible that we have already been utilizing PaLM without even realizing it. Notably, PaLM is not limited to translating spoken languages but can also handle programming languages.

The versatility of PaLM is exemplified by its ability to facilitate collaboration between individuals with different language backgrounds. For instance, an English-speaking individual can use PaLM to collaborate with a colleague in a codebase where all the documentation is written in Korean. Moreover, PaLM excels in generating and comprehending nuanced language, including idioms and riddles. This capability is crucial as it requires understanding not only the figurative meaning of words but also the literal intent behind them.

As LLMs continue to evolve and be integrated into various applications, we can expect further advancements in language processing and understanding, benefiting individuals and organizations across diverse linguistic and professional contexts.

PaLM 2 is a language model developed by Google that aims to bring AI smarts to some of Google's most popular apps, such as Gmail, Google Docs and Bard.

Music Language Model (MusicLM)

"Musik kann die Welt verändern" - Ludwig Beethoven

As Ludwig van Beethoven famously stated, which translates to "Music can change the world." When it comes to music generation, there is indeed a remarkable potential for transformation. One notable model in this realm is MusicLM, which has the ability to generate music from text captions without utilizing any diffusion techniques. What sets MusicLM apart is its ability to consistently generate music at 24 kHz, maintaining a high level of fidelity over several minutes.

MusicLM is a cutting-edge audio model that operates using two types of tokens: semantic and acoustic tokens. The semantic tokens capture the essence of melody and rhythm, while the acoustic tokens, provide valuable information about the recording conditions of a track. These tokens come together in a remarkable fusion, resulting in the creation of high-quality audio tracks. One of the most impressive features of MusicLM lies in its melody conditioning capabilities. This means it can generate music that precisely matches a given text prompt while incorporating the desired melody seamlessly. What's even more astonishing is that MusicLM has been put to the test with descriptions of paintings, showcasing its remarkable ability to translate the mood and atmosphere of visual art into mesmerizing musical compositions.

Another intriguing feature of MusicLM is its generational diversity. The same text prompt can yield a range of music compositions or variations within the same sample. This flexibility opens up countless creative possibilities for musicians and composers. AI has the potential to revolutionize music production in the future, envisioning a scenario where a smart keyboard can generate numerous arrangements with different instruments, sounds, and styles based on a simple melody played by the musician. Instead of engaging in the laborious process of composing and producing music from scratch, artists can guide AI systems to create captivating tracks until the desired result is achieved.

The integration of continuation and voice cloning techniques further expands the horizons of music production. Imagine the ability to bring legendary artists like Michael Jackson and Kurt Cobain back to life, enabling them to sing a duet on your track using their cloned voices. Additionally, leveraging ChatGPT, you can employ AI assistance to craft the lyrics for your compositions. To complete the immersive musical experience, Vision AI can be utilized to capture the mood and emotions displayed on the faces of the audience, allowing real-time generation of music that perfectly suits their reactions and preferences.

As advancements continue, musicians, producers, and listeners can expect a groundbreaking era of creativity, collaboration, and personalized musical experiences. The power of music to shape the world is undeniable, and AI's pivotal role in its evolution promises a thrilling and harmonious journey ahead.

MusicLM can create high quality music from text captions

.

Conclusion

In conclusion, the emergence of Generative AI marks an exhilarating new era filled with endless creativity and boundless potential. From the transformative power of Large Language Models like PaLM to the enchanting melodies crafted by MusicLM, AI systems are reshaping industries and revolutionizing artistic expression. As we continue to push the boundaries of what AI can achieve, the possibilities are thrilling but uncertain. As we explore the vast potential of generative AI, it becomes imperative to confront the complexities of responsible deployment. Our creative endeavours must uphold the core values of privacy, fairness, and inclusivity, guiding us towards an AI-driven entertainment landscape that reflects the best of humanity's ingenuity and empathy. The future of generative AI is a symphony of human ingenuity and machine intelligence, harmonizing to create a world where imagination knows no bounds. By embracing AI's potential with a mindful approach, we can shape a future where creativity flourishes while safeguarding against unintended consequences. Let us embark on this journey with optimism, curiosity, and a deep commitment to using AI ethically and responsibly, shaping a world where art, music, and beyond thrive with the best of human and machine collaboration.

"The real problem is not whether machines think but whether men do"
- B.F. Skinner

Music Room in Graceland, the home of Elvis Presley

UDAI SHANKAR

Director- Data Science, Providence

Udai Shankar is a data science leader and researcher having 18+ years of experience in solving real-world business problems in various domains including operational intelligence, NLP and security. Udai has previously held senior roles at CA Technologies. He has strong knowledge of algorithms with experience in understanding and shaping use cases.

To start can you share about your personal journey in data science and machine learning?

I started learning data science sometime in 2011. Prior to that, I was a C++ developer with CA technologies, and we've done some interesting work on compilers and a lot of good stuff on C++ core programming stuff. Then in 2011 I took a course in pattern recognition from IIIT Hyderabad, and that's how I got started. Around 2013 I shifted totally into data science with CA technologies and joined Providence in 2019.

Your interests also lie in topological data analysis and quantum machine learning. These are very niche topics involving mathematics and computer science. What inspired you to pursue research in those specialized areas?

I love mathematics, that's what got me interested in machine learning in the first place. I studied as a part-time student at IIIT Hyderabad. During this time, I found quantum computing very interesting. I studied theoretical computer science and then quantum computing at IIIT itself. I am an independent researcher in Quantum Algorithms, more specifically, I study the homology groups that arise in topological data analysis from the lens of quantum hidden subgroup problem.

That's fascinating. So I myself come from a mathematical background from CMI. We had algebraic topology in our curriculum in both undergrad and master's. In the theory of homology, wherein we can count the number of holes in some topological spaces. I want to know how this whole detection formalism of algebraic topology can be connected with these robust computational methods to extract qualitative features in data.

The area that I research is an intersection of computational complexity, quantum hidden subgroup algorithms and computational topology. Since you are into math, you might have studied, algebraic topology, which cannot be learned without the algebraic fundamentals – basically abstract algebra (groups, rings and modules). There is a result in quantum computation that says that if there is an abelian group, it is easy for a quantum computer to compute the generators of a hidden subgroup in that abelian group. This result is important because all the quantum algorithms where an exponential speedup was observed can be reduced to instances of abelian hidden subgroup problem (AHSP). For example the Shor's algorithm, discrete log, period finding etc., are all instances of AHSP. There are also some results when the groups are not abelian [like Normal subgroup, dihedral subgroup etc.]. Now, (free) abelian groups arise naturally when we work with homology. Topological data analysis (TDA) links data and topology (persistent homology etc.). So I study the groups that arise in TDA from a quantum lens. There is one famous result on quantum algorithms for TDA by Seth Lloyd, but this is not from the algebraic perspective (i.e., not directly using AHSP).

So the way I entered into this is just out of curiosity and exploration. Math excites me, so machine learning excites me, and this also excites me. Topological data analysis is very powerful. Not the best out-of-the-box algorithm, but probably the internals of it can be used in your day-to-day data science as well. That's my perspective.

"Math excites me, so Machine Learning excites me"

Okay, that's actually nice to hear because, from what we have seen in theoretical sciences, it's difficult to connect with reality if you go into those niche topics. So are there any applications of topological data analysis that have been going on in the industry?

There is an entire company that spun out of topological data analysis called Ayasdi. They started with topological data analysis. There is one problem with topological data analysis. Our edge node graphs are one-dimensional topological objects. In machine learning, we have probabilistic graphs where each node is represented with a probability distribution, so that's how we study machine learning, isn't it? Everything in machine learning is about an implicit graph & (implicit) probability distribution on the nodes, and once, we have the probability distribution, we are trying to predict something using the probability distribution. So basically, under every scenario, we have a probabilistic graph. Now, think about a scenario where blood pressure is, let's say, one node and diabetes is another node. We might have an edge between them, that is a 1D graph. Suppose, we ask, how do blood pressure and diabetes function together? When a person has both diabetes and blood pressure, what is the curve that determines their, let's say, disease progression? In other words, if I am studying BP, Diabetes and the future state of CKD together, I have to take not just the edge between diabetes and blood pressure, I have to take all three together, and that's a simplex. So when you take three nodes together, then the entire triangle is a simplex. That's how higher-dimensional topological objects enter into the picture. An interesting area to explore is persistent homology.

AI algorithms have already matched the performance of human experts on several prediction tasks. But humans still have some valuable domain knowledge, which hasn't been incorporated into the learning process. In this context, how do healthcare professionals collaborate with AI systems to combine human expertise in decision-making?

I'll answer this from first a theoretical perspective and then come to the practical aspects of it. So, from the theoretical perspective, there is science, and then we have all our machine-learning algorithms that can predict, cluster etc. Suppose you're trying to do a controlled experiment, where you want to see whether a particular drug really affects a particular disease. The domain knowledge must be passed as an input to the algorithm because domain knowledge is not available in data, so it has to be passed as an input to the algorithm. Data can also be used to validate different domain perspectives. Think about it this way. How do I present my domain knowledge to the algorithm? It must be a mathematical object. So, my take on this is, that the best way to present domain knowledge to an ML algorithm is to provide a causal graph [or a higher dimensional topological object]. Your algorithms work on the causal graph with the support of the data.

As an extension of the previous question, how do you see the role of Bayesian models in healthcare AI and how do they differ from traditional machine learning models? Could you give some examples of healthcare applications where Bayesian models have really shown advantages?

The main power of Bayesian modelling is creating custom distributions and inferring them using MCMC or variational inference, right? For example, let's say, model a coin. I know that coin can be modelled as a Bernoulli distribution. Now the question really is, suppose I want to model a different type of coin – maybe a coin with a memory. That means it remembers its previous toss. Suppose I claim I created this coin and give it to you. Now model this coin in terms of a probability distribution. It's way easier and clearer in a Bayesian setting than in a non-Bayesian setting. Also, the causal graphs and all that I was talking about, can be easily correlated or linked to a Bayesian setting than a non–Bayesian setting. If I know that F causes A, I can directly put a distribution on F and A and then link them together, and the problem becomes extremely simple to solve. So, Bayesian mod-

"Science is about discovering the causal links , I propose a causal link and verify whether it is causing it or not"

"Bayesian modelling is very, very fundamental. It is providing you with the tools to model real problems"

elling is very, very fundamental. It provides you with tools to model real-world problems. Suppose I were to give the same information to a deep learning algorithm without any Bayesian modelling [Of course, Bayesian deep learning & Graph NNs exist, but for the moment, imagine we are not using any of it] I'm actually not telling the algorithm anything about the causal links (the Science or the Domain) or about the probabilistic graph that is available from the domain. The causal structure is better represented naturally through Bayesian learning. So, for non-Bayesian approaches, you have to do something unnatural about it.

Now that we have a lot of talk about data privacy and there is a Data Protection bill in India. So in the context of healthcare, how do you handle sensitive or confidential patient data while communicating visualizations to your stakeholders?

Yeah, so there are two levels to it. One is where we work within the internal structure of providence, where we have access to data. Before I use data for a specific purpose, there's a governance board known as the IRB Board whose permission we take and then work on our research project. From an external perspective, when I want to give out data to, let's say, a partner. We de-identify the data (HIPPA Compliant data). We make sure that all the PHI and PII data is totally masked (or obfuscated), and we change dates in such a way that the deltas remain the same [so that ML is still possible on the data]. We also mask out the zip codes, which have less than, let's say, 20,000 people. The ages are converted into buckets so that age groups with few people cannot be used to identify the person. These are the kinds of things we do for privacy. All the data that goes to any algorithm usually goes through this de-identification process. We have a service that does de-identification on the data.

As a follow-up question, is there any regulatory compliance which you have to follow in case you have to take data of US patients and work in India?

It's just the same, we go through the IRB process, as I said, which is the internal regulatory board. Providence India is just a part of Providence overall, so there is no separation or anything out there.

As you earlier said, like all of us are hooked to ChatGPT and large language models these days. So in light of the availability of advanced language models such as ChatGPT, modern chatbots are being used to enhance customer experience in the case of healthcare. Considering there have been instances where chatGPT is giving out wrong answers and giving out inaccurate information, how is the healthcare industry making sure that it can leverage these models to its advantage?

We obviously have to finetune (or retrain) these models for our purposes. As it is, the hugging face LLM models are not very useful to us because they are huge and because of the high computational requirements. Also, the open-source LLM models are not trained for the specific use cases we are interested in. Let's say you are a doctor and you have a specific question in your mind and you come to the clinical data (EHR data) to gather some output as to what the data is saying about your Hypothesis (the specific causal question in your mind). So you've got a causal link that you want to validate through the data. That's one type of question. The other type of question is, let's say, for example, how many people are coming to the emergency department, i.e., the footfall at the emergency department, and how

" Chat-GPT is not a plug-in, but the inherent technology is very useful still. You know, there is a dialogue element to it, a human feedback element to it. "

to schedule nurses and things like that.

Let's talk about yet another kind of problem where I want to predict the revenue of the company. So I want to look at how the revenue is getting generated and where I can improve and get more revenue and all that. That's another kind of scenario. None of these scenarios are directly amenable to working with the LLMs. The data is usually a collection of disparate pieces of structured and unstructured information that need to be combined and linked together in a systematic way in order to run any deep learning algorithm on it.

You were talking about, extending this to images and genomics data, for LLMs. Could you talk about what exactly you do there when you're using images or genomics data for LLMs?

We have not started working on the images and genomics yet. Even with the EHR data, there are two parallel worlds if you think about it. The unstructured data is one world of information, for example, patient notes - the doctor has written some notes, and the images have their diagnostic reports. Then you have this structured data, which includes things like the BP measurement or, let's say, their heart rate, and pulse rate. The structured data also includes the medication, diagnosis etc. The first problem really is how do we bring these two worlds together. My claim is through discretization, vectorization and graph NNs. For example, we can build vectorial representation from textual data on structured and unstructured data. We can bring out a connection.

between words that are close by. Then I can draw an edge between them. So, I made it into a graph, thus shifting from the textual to the graphical picture. If we want to do a proper analysis of the data, our LLMs have to work over a causal structure derived from data and domain knowledge.

From your answer, it was clear that ChatGPT can't be the plug-in, plug-out solution for everything. Is it?

Exactly. It is not a plug-in, but the inherent technology is still very useful. The backend, which is, Transformers + RLHF is extremely powerful. But this has to be made to work over the data specific to our domains and with an understanding of the causal structures inherent in my problem.

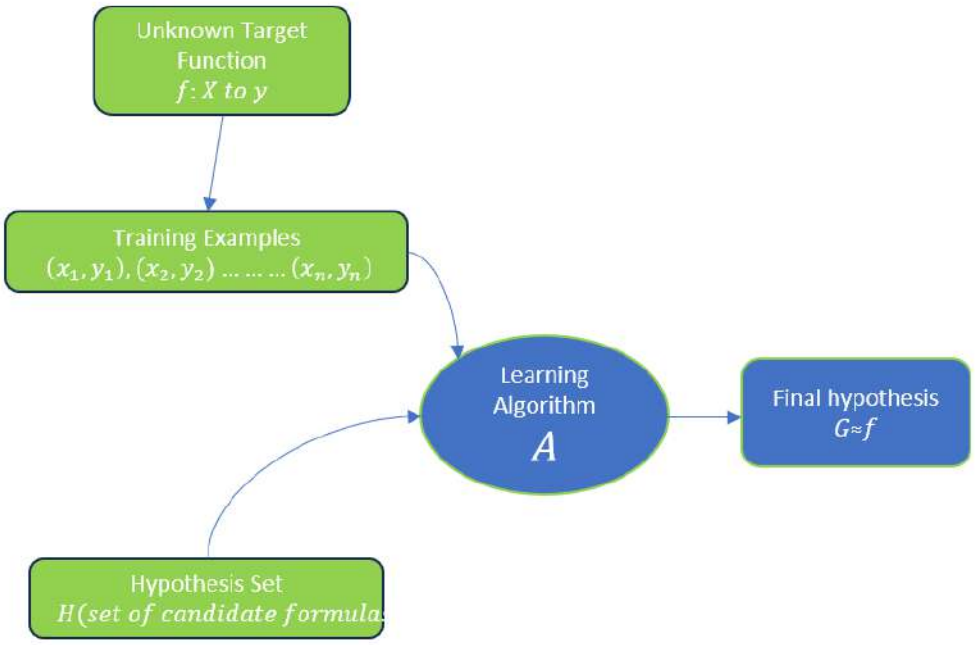

Synthetic Data Generation
Robust Modelling with Limited Data
Parijat Pushpam

"God made the integers; all the rest is the work of a man"- Leopold Kronecker

General overview / Setting the stage

At the heart of every Machine Learning model lies the fundamental concept of "learning from data." This singular essence sets Machine Learning apart from conventional programming and ignites our fascination with its potential to revolutionize industries and shape the future. The question is, how do we carry this out? Before answering this question, we should look at a famous scheme or flow of things that are, available in the data science literature everywhere.

This is how we get to find a relationship between the input and the output. But how are we going to make sure that our model would perform just fine when it encounters new data? Now, this right here is the end goal any machine learning model wants to achieve. "To minimize the generalization error." So, the essence of machine learning is - Train your model with observed data and try to minimize the error on data which you have never seen.



The mathematical notations and symbols may ward off a significant portion of the readers. But, sometimes hanging onto things tends to reward us in unimaginable ways. Like Werner Heisenberg once said: "The first gulp from the glass of natural sciences will turn you into an atheist, but at the bottom of the glass God is waiting for you."

In pure layman's terms, what we do in machine learning is to find a function that closely approximates how the given data behaves. There could be multiple functions through which we could model our data, but there must be some jackpot function that mirrors our data efficiently. We get to this efficiency using some kind of error function that punishes the algorithm for making errors. And voila!! That is how we pin down a model of the given data.

Now what if the data for doing the above sorcery is unavailable in the first place? What if the data collection is dangerous or what if the data does not even exist because you are just hypothesizing a model or what if there are cost constraints (A labelling service could cost $6 to label a single image)? What if the unavailability is due to privacy concerns? Nobody wants to disclose sensitive or private information and that too on a large scale.

It seems that we have hit a roadblock and all the fast-paced progress needs to chill out. Enter Synthetic Data Generation. "Data is the new oil," we have been told 'n' number of times (n tending to infinity). But what we fail to realize is that very few are sitting on the gusher, many are making up their own oil.

What is Synthetic data and why do we care about it?

Synthetic data generation techniques involve creating data through processes such as physics-based simulations, procedural generation, or data augmentation. These techniques allow the generation of synthetic images, videos, text, sensor data, or any other relevant data type to the specific domain or application. Synthetic data is annotated information that computer systems or some algorithms generate as an alternative to real-world data, sometimes even better than the original one. In even simpler terms, it is created in the virtual world rather than being collected or measured from the real world via painstaking experiments. Statistically, this generated data mimics the real world and the difference between them is often dependent on the algorithm or method used for their creation. Often synthetic data has certain advantages over collected data. Deep learning models require tons of data to train efficiently and synthetic data would ease the process. Cost cutting is just one part of it, we could even include rare or corner cases in our dataset which amounts to better diversification. In addition to this, privacy issues would barely surface. Not only this, it would enhance the research process and minimize the time required to collect data, thereby giving more time to devote to actual research purposes.

Not that it is a new thing, synthetic data and its generation has been there for ages. Be it simulated driving or flying in video games or scientific simulations of fundamental particles of nature.

Current Usages

Amazon

Currently, **Nvidia's Omniverse replicator** generates 3D synthetic data to train autonomous vehicles to navigate safely amid shopping carts and pedestrians in a simulated parking lot. For training autonomous vehicles, we need huge datasets and multiple situations cannot be recreated in real life due to the cost restrictions. The use of synthetic data allows **Amazon Robotics** to create a large and diverse training dataset, providing ample opportunities for the robots to learn from various package scenarios. This approach enhances the robots' object recognition capabilities and helps them perform effectively in real-world logistics operations, where they encounter a wide range of package types and sizes.

Healthcare

Curai - a US-based healthcare startup leverages state-of-the-art synthetic data generation techniques to obtain simulated images and train its diagnostic neural network model to deliver better diagnosis using its text-based virtual agent. Although they did not disclose their data generation techniques anywhere, GANs may have assisted them in that pursuit. We will be getting a taste of how GANs work afterwards.

Fraud/Finance

American Express uses synthetic data generated by GANs to improve its credit fraud detection models due to the scarcity of real fraudulent transactions in the training data. This synthetic data helps the model learn rare fraud patterns more effectively.

Synthetic Data is used to train fraud detection models for credit cards

Startups

As many as 100 startups have surfaced specializing in providing artificial and domain-specific data each with their own secret sauce. Take the case of **Nuvanitic,** which specializes in creating synthetic data in clinical trials to facilitate medical innovation in drugs and devices. Another exciting company **FinCrime Dynamics** allows financial institutions to create their own synthetic datasets enriched with customized financial crime simulations. It also provides a safer way for financial institutions to understand and improve the performance of their financial crime controls.

How synthetic data generation works

Intuition tells us that if we have a vector of parameters and somehow know how the distribution function will behave (probably domain knowledge), then we have probability values. Almost the same approach is used in methods currently practised.

Domain Randomization

The purpose of domain randomization is to create a diverse and varied training set, which covers a wide range of possible variations. By randomizing the parameters, we ensure that the synthetic data samples different configurations and scenarios, allowing the machine learning model to learn robust and generalized representations. Domain randomization does not necessarily require preexisting data points. In fact, one of the advantages of domain randomization is that it allows us to generate synthetic data without relying on real-world data. Instead, we define the domain and its properties and then randomize those properties to create synthetic data.

Lets denote the synthetic data as χ and the randomized parameters as **P.** The rendered synthetic data **rnd** is obtained by simulating the domain **D** with the randomized parameters **P**. This process can be shown as χ_{rnd} = **f (D, P)**, where '**f**' is the rendering function. If supervised learning is involved, the synthetic data χ_{rnd} is labeled with ground truth information **Y.**

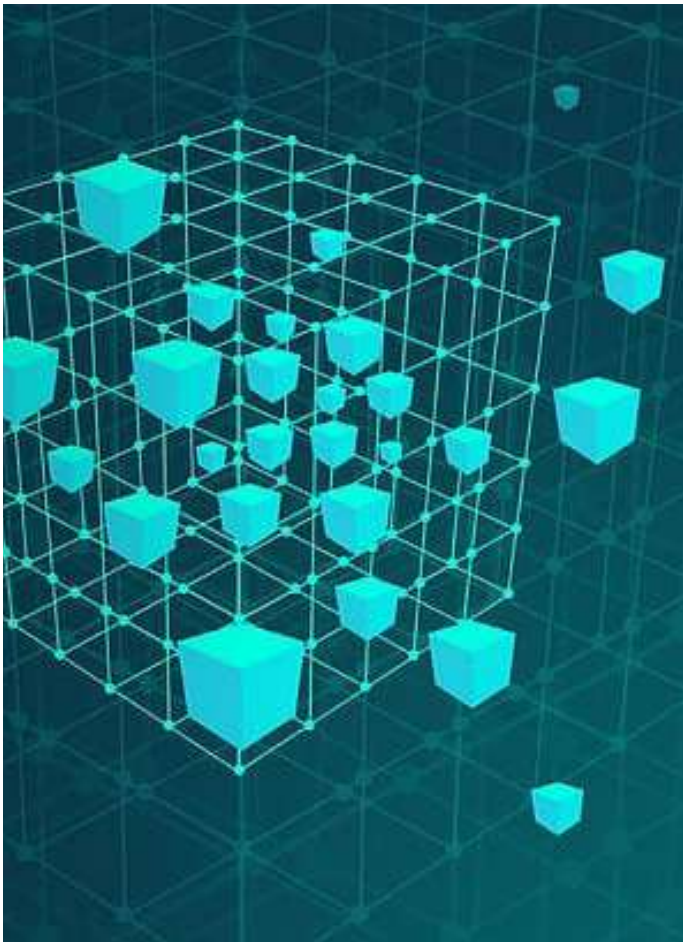
Take the case of the generation of synthetic data for autonomous vehicle simulation -

The domain D in this example represents a virtual environment that simulates real-world driving scenarios. It includes elements such as road layouts, traffic signs, pedestrians, weather conditions, and other dynamic elements.

The randomized parameters P refer to the set of variables that introduce variability and randomness into the simulated environment. These parameters control factors like weather conditions (e.g., sunny, rainy, foggy), traffic density, road surface conditions, and vehicle dynamics (e.g., different car models, speeds, and behaviours).

The synthetic data χ is the output of the simulation process, which consists of various data types, such as images, sensor readings, and other relevant information, representing the virtual environment and its dy namic elements. For instance, χ could include images captured by virtual cameras mounted on virtual vehicles, lidar point clouds, GPS coordinates, and so on.

Domainrandomizationhelpsincreatingadiverse training dataset



Generative Adversarial Networks

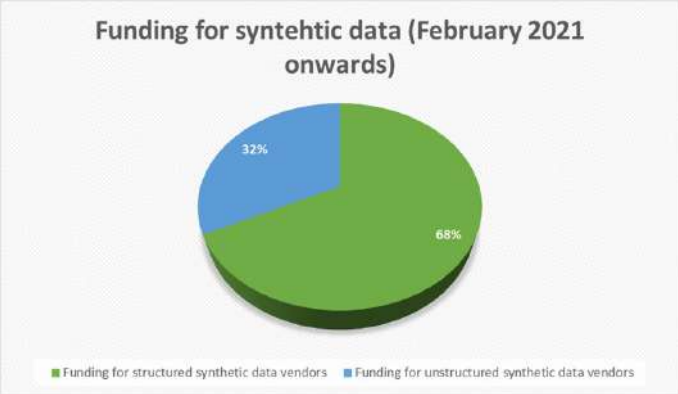
It is a class of deep learning methods used to generate synthetic data which resembles a given dataset. The network consists of a Generator and a Discriminator which functions as its main components. The generator learns to generate synthetic data samples that mimic the given dataset while the discriminator makes sure that it catches the fake ones until it can no longer differentiate between real and fake ones.

This tussle between the discriminator and its counterpart takes place during the training of GANs in the background, while we look at the computer screen with a nonchalant attitude and sip our coffee.

The feeling that during its initial training phase, the network would do very badly. That feeling is quite right having stemmed from the fact that the initial input is just a random noise, and catching it as fake is just a child's play for the network. But as time passes, the loss functions will tell the network that – 'Hey, you are doing bad, it is high time you step up'. This will make the network adjust a few parameters to the point that discrimination between fake and real data points is barely possible.

GAN's are models that can synthesize data that is similar to a given dataset

The Future of synthetic data generation and its business implications



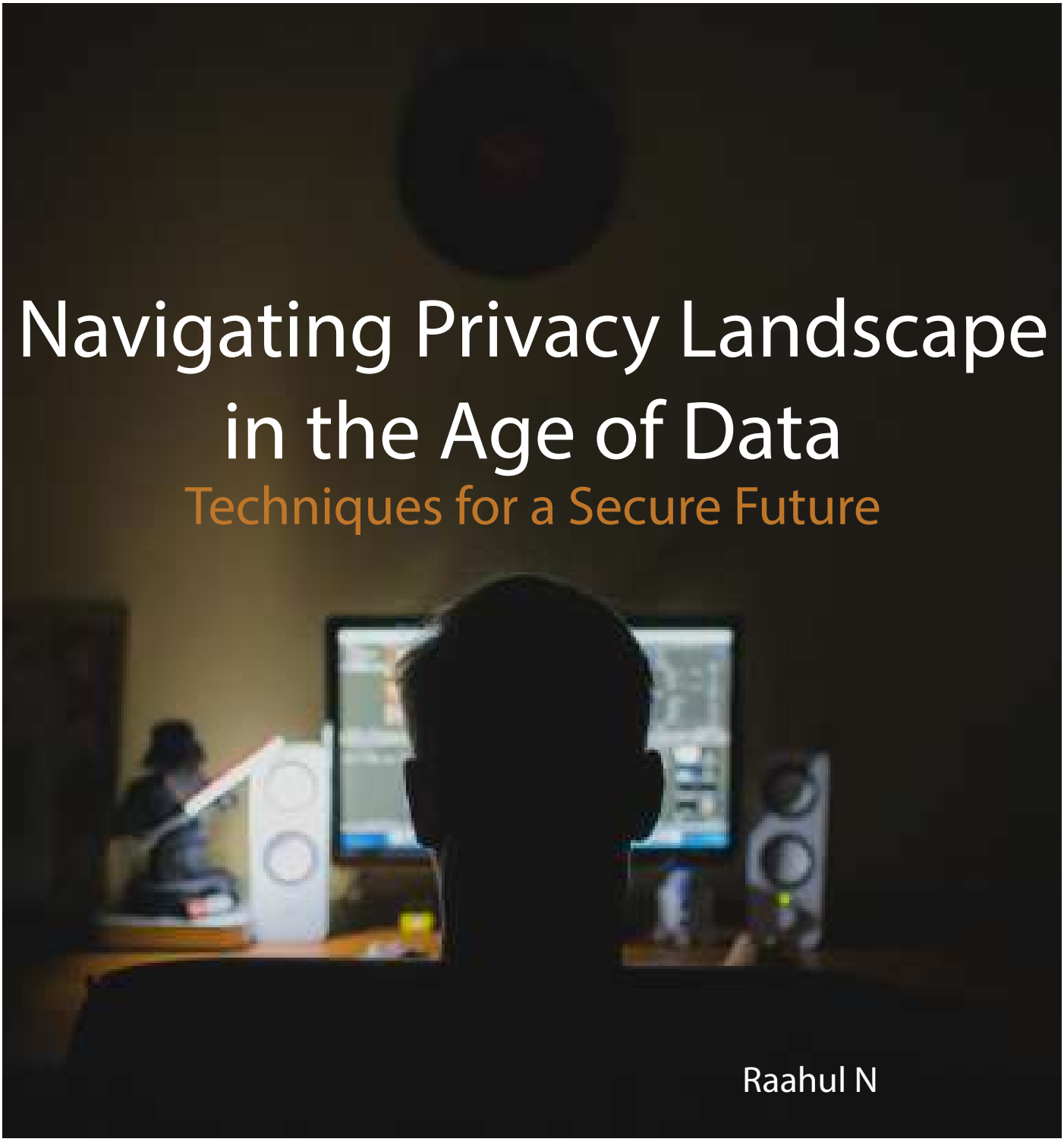
The synthetic test data market is experiencing rapid growth, with numerous vendors offering solutions in this space. However, there is often confusion regarding the terminology used by these vendors. Some vendors provide rule-based simulated data, while others specialize in AI-generated synthetic data. Despite the differences, the overall technology is gaining traction, partly due to fewer privacy constraints associated with developing fake data, we have sites like Amazon & Walmart selling curated data, apart from their conventional offerings.

As many as 100 companies that aim at commersializing synthetic data and products for multiple purposes have surfaced. No doubt, sometime in the near future, we have sites like Amazon & Walmart selling curated data, apart from their conventional offerings.

Amazon's AWS has come up with DeepRacer which is primarily used as a platform for learning and experimenting with reinforcement learning and autonomous driving algorithms. The main goal of DeepRacer is to provide developers and enthusiasts with a hands-on experience in building and training autonomous driving models using AWS services. Users can train their DeepRacer models using the AWS DeepRacer console, which provides a web-based interface for managing and training the models. In the context of DeepRacer, synthetic data generation can be used to augment the available training data.

By using DeepRacer's cloud-based training and deployment infrastructure, businesses can significantly reduce the time required to train and optimize autonomous driving models. This accelerated development cycle allows for quick prototyping and experimentation, ultimately giving rise to faster iterations and improvements.

Amazon introduced Deepracer to experiment with autonomous driving algorithms

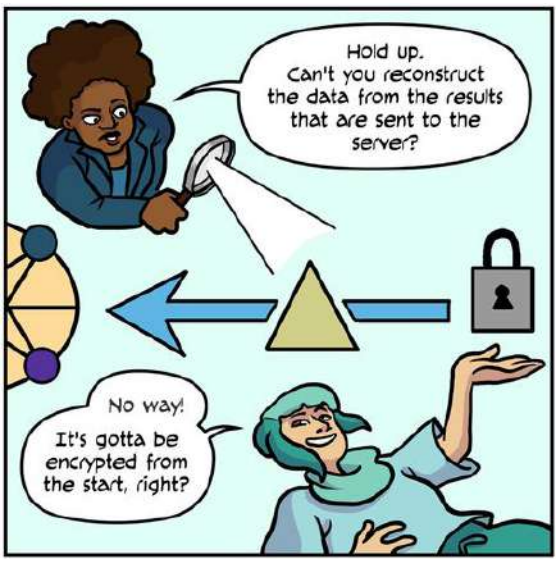


"Arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say"

- Edward Snowden

It was early June 2013. Edward Snowden, a former NSA(US) staff, revealed to the world about the surveillance programs of the U.S. Government. The classified NSA documents leaked by Snowden brought the topic of "problematic surveillance programs" to international attention through journalists.

Around the same time, a British consulting firm (Cambridge Analytica) collected about 87 million Facebook users' personal data via Meta's Open Graph Algorithm & built psychological profiles using specific questionnaires. These profiles were allegedly used for analytical assistance in US presidential campaigns in 2016. The same firm was also widely accused of interfering with the UK's Brexit referendum. These 2 major events brought to public awareness the extent of psychological targeting using personal data.



If you are an individual or an organisation working with your users' data for insights or any interested common citizen, this article aims to guide you through some privacy-preserving techniques in machine learning with a 10,000 ft overview of some of their working principles.

Specifically, we'll be going through Differential Privacy, Secure Multiparty Computation, Federated Learning and Homomorphic encryption.

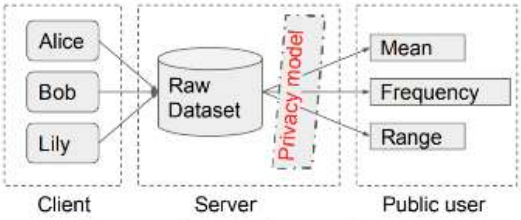
Differential Privacy

One of the most widely used techniques for privacy-preserving in machine learning is differential privacy. Differential privacy is a mathematical framework that provides strong guarantees of privacy by adding a controlled amount of noise to the data before it is analysed. This helps to ensure that the output of the analysis does not reveal any sensitive information about individuals who contributed the data.

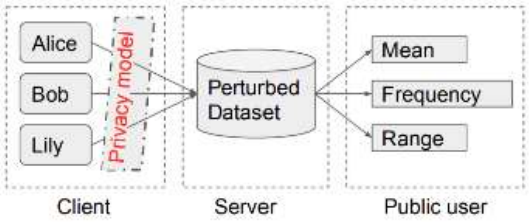
In the Indian context, the Digital Personal Data Protection Bill (withdrawn now owing to recommendations received through public consulting) is in the works and we can expect it in the coming years. These techniques include practices such as data anonymisation, encryption, which help protect sensitive information while still allowing for analysis and insights. These techniques enable organisations to strike a balance between utilising data for valuable insights and maintaining the privacy and rights of individuals, thereby promoting a responsible and ethical data-driven environment.

Since the Cambridge Analytica incident, controversy centre Facebook (now Meta) has always been under the radar for privacy violations & targeted ads. The firm has been fined billions (with a b) of dollars for numerous data breaches & privacy violations. These incidents in the last decade sparked public interest in online privacy and dangers of misuse of personal data. Ever since big techs & governments across the world are striving to meet the people's need for their privacy.

Frameworks like GDPR were set in place by governments to protect individuals from being exploited with their personal data. For people new to the 'Online Privacy' scene, General Data Protection Regulation (GDPR) is a comprehensive privacy law brought forth by the UK Govt. that sets guidelines for the collection, processing, and storage of personal data of individuals. The aim is to protect the privacy and rights of individuals by giving them control over their personal information. As a result, organisations are required to implement privacy-preserving techniques to ensure compliance with GDPR and safeguard individuals.



(a) Centralized differential privacy



(b) Local differential privacy

The concept of differential privacy is based on the idea of adding randomness to the data such that the analysis cannot be used to identify individual participants. However, the analysing algorithm is still able to produce statistically meaningful results without compromising the privacy of the data subjects. One intuition for this would be to think about aggregates. A mean of a large data, with random noise added to individual values would still remain the same. Some popular techniques that use differential privacy include randomised response, local differential privacy, and global differential privacy. One example of an application of this is in medical data. Differential privacy enables the analysis of sensitive patient data, electronic health records, and genomic data while maintaining patient privacy.

Differentialprivacyaddsrandomness to data to avoid user identification

While global differential privacy model collects all the users' data and releases the perturbed version. This poses a privacy risk, as time and again we have seen, that any single data curator cannot be trusted. Local differential privacy solves this by perturbing the data before it leaves the device. Now only the owner of the data can access the original data, which provides much stronger privacy protection for the user.

If we take machine learning into account, this involves incorporating the concept of differential privacy into the training process of the ML model. One approach involves adding noise to the gradient of the loss function (differentially private stochastic gradient descent or DP-SGD) during the training process.

$$\theta_{t+1} \leftarrow \theta_t - \eta \cdot (\nabla_t + b_t).$$

Here the model parameters (θ) are updated by subtracting this gradient multiplied by a small constant (η). Gaussian noise (bt) is added to their sum to obtain the indistinguishability needed for Differential Privacy. This helps to ensure that the updates to the model are not overly influenced by any one individual's data. Another approach involves using a differentially private data synthesiser to generate a synthetic dataset that can be used to train the model.

Secure Multiparty Computation

Secure multiparty computation (MPC), as the name suggests, is a cryptographic technique that involves multiple parties jointly computing a function on their private inputs without revealing their individual inputs to each other. The goal is to preserve privacy and confidentiality while obtaining the desired computation results.

MPC achieves this by utilising cryptographic protocols and techniques such as secret sharing, secure function evaluation, and secure two-party computation. Here's a high-level explanation of the technique:

1. Secret Sharing: The private inputs of each party are divided into shares using a secret sharing scheme. Each party holds a share of their own input and does not have complete knowledge of the inputs of other parties.

2. Secure Function Evaluation: The parties then perform a series of computations on their shares while exchanging information in a secure manner. This involves executing cryptographic protocols that allow the parties to compute intermediate results without revealing their individual inputs.

3. Secure Two-Party Computation: In some cases, MPC involves only two parties. Secure two-party computation protocols ensure that the computation can be performed while preserving privacy and confidentiality. These protocols utilise cryptographic techniques such as garbled circuits, oblivious transfer and zero-knowledge proofs.

4. Consistent Output: Through the secure computation process, each party learns the final result of the joint computation without gaining any information about the other party's private input. The output is consistent with what would have been obtained if the computation had been performed on the combined inputs directly.

For example, if we are computing the average height of three people (Ayesha, Bindu and Catherine) without revealing individual heights.

Secure multiparty computation involves multiple parties jointly computing a function on their private inputs

• Say Ayesha's height is 140cm. Here, 140cm is split into 144cm, -11cm and 7cm

• Ayesha keeps one of the 3 pieces with herself and shares the other two to others.

• The same steps are followed by Bindu & Catherine.

Ayesha	Bindu	Catherine	
144	-11	7	140cm
-6	132	24	150cm
20	0	140	160cm
158	121	171	

Sum of their encrypted pieces
= (158+121+171) = 450cm

Average height = 450/3 = 150cm

Federated Learning

Federated learning is a privacy-preserving approach in machine learning where data is kept decentralised and computation is performed locally. Instead of sending raw data to a central server, federated learning allows devices or entities to collaboratively learn from their respective data without sharing it directly.

Here's how it works: The learning process takes place on individual devices or edge nodes, such as smartphones or IoT devices, which possess their own local datasets. These devices train a model using their local data, and instead of sending the data itself, they send model updates or gradients to a central server.

The central server then aggregates these updates from multiple devices to create a global model that represents the collective knowledge learned from all participants. The updated global model is then sent back to each device, allowing them to improve their local models based on the collective insights.

Federated learning provides several advantages. Firstly, it enhances privacy since the raw data remains on the local devices and is not shared directly with the central server. This mitigates concerns about data breaches and unauthorised access. Secondly, it enables collaboration on a large scale, allowing diverse entities to contribute their knowledge without having to centralise the data. Lastly, it promotes efficiency by reducing the need for large-scale data transfers, especially in scenarios where data size or connectivity is limited. One of the advantages of federated learning is that it allows the model to be trained on a more diverse dataset, which can improve the accuracy of the model. It also reduces the risk of a data breach, as the data never leaves the local device. Some examples of federated learning applications include mobile keyboard prediction and medical image analysis.

Homographic Encryption

Homomorphic encryption is a remarkable technique in cryptography that allows computation on encrypted data without needing to decrypt it first. This means that you can add, subtract, multiply, and divide encrypted data without knowing what the unencrypted data is. Homomorphic encryption makes it possible to train a machine-learning model on encrypted medical records without reading the actual records.

Homographicencryptionallowscomputationon encrypteddatawithoutneedingtodecryptitfirst

Let's say you have two encrypted numbers, and you want to add them together. With homomorphic encryption, you can perform the addition operation directly on the encrypted data, yielding the encrypted result. No one can peek inside the box and see the actual numbers or the result, but you still get the correct result!

Wide range of potential applications of homomorphic encryption include:

- E-voting
- Healthcare data sharing
- Financial transactions

Homomorphic encryption is still a relatively new technology, and there are some challenges that need to be addressed before it can be widely adopted.

One challenge is that homomorphic encryption is computationally expensive. Another challenge is that homomorphic encryption is not yet as secure as traditional encryption methods. Despite these challenges, homomorphic encryption is a promising technology with the potential to revolutionise the way we protect and use data.

Privacy-preserving techniques and technologies are becoming increasingly important in machine learning and data science. Differential privacy, secure multiparty computation, federated learning, and differential privacy for machine learning are just a few of the techniques that are being used to address these concerns. As the field continues to evolve, it is likely that we will see new and innovative techniques emerge that further enhance privacy preservation in machine learning and data science.

As mentioned earlier by embracing these techniques, we can strike a balance between extracting valuable insights from data and upholding the privacy rights of individuals. Let us remain vigilant, adapt to emerging techniques, and collectively work towards a future where privacy and data-driven innovation coexist harmoniously.

THE TRI-INSTITUTE ANALYTICS SUMMIT

TRILYTICS

600 + PARTICIPATING INSTITUTES

5200 + REGISTRATIONS ON unstop

in association with

World Wide Technology

The second edition of Trilytics, the prestigious annual analytics summit organized by the Post Graduate Diploma in Business Analytics (PGDBA) students, successfully concluded at the Indian Institute of Management Calcutta (IIM Calcutta) in Joka. The event was supported by lead sponsor World Wide Technology (WWT), associate sponsor State Bank of India, and media partner Times of India. Drawing an impressive global audience of over 5200+ participants, the event's analytics case competition, held in collaboration with WWT, brought forth innovative solutions to real-world challenges. Distinguished industry voices, including Ajay Dadheech, Head of Management Consulting APAC at WWT, and Udai Shankar, Director of Data Science at Providence, led insightful keynotes.

Inaguration session by Chairperson of PGDBA Dr. Samit Paul

Ajay Dadheech, Head of Management Consulting APAC at WWT

Udai Shankar, Director of Data Science Providence

THE TRI-INSTITUTE ANALYTICS SUMMIT

A notable panel discussion centred on the trending topic of Generative AI and its responsible applications. Distinguished panellists included Dr Adway Mitra from the Center for Excellence in AI at IIT Kharagpur, Achal Sharma, Lead Data Scientist at WWT, and Venkat Subramanian, Director of Data Science at Paypal. Prof. Dr Sowmyakanti Chakraborthy adroitly moderated the panel discourse from IIM Calcutta. The pinnacle of the event was the analytics case competition, crowning Team RVB from the Indian School of Business as the victors, followed closely by teams from PGDBA and IIM Bodh Gaya.The inaugural ceremony witnessed enlightening speeches from Prof. Virendra Kumar Tewari, Director of IIT Kharagpur, Prof. K Sudhakar Reddy, Dean of VGSOM, and Prof. Samit Paul, Chairperson of PGDBA, setting a tone of inspiration for the engaging sessions that followed.

Left to right: Mr. Achal Sharma (WWT), Mr. Venkat Subramanian (PayPal), Dr Adway Mitra (IIT Kharagapur), Prof. Dr Sowmyakanti Chakraborthy (IIM Calcutta)

References

Towards Artficial Intelligence: Recent Developments and Potential Risks

- https://blogs.windows.com/windowsdeveloper/2023/05/23/bringing-the-power-of-ai-to-windows-11-unlocking-a-new-era-of-productivity-for-customers-and-developers-with-windows-copilot-and-dev-home/
- https://www.adept.ai/
- https://openai.com/customer-stories/be-my-eyes
- https://www.prnewswire.com/news-releases/humanoid-robot-market-worth-17-3-billion-by-2027--exclusive-report-by-marketsandmarkets-301520783.html
- https://indianexpress.com/article/technology/artificial-intelligence/openai-backed-robotics-firm-1x-deploys-the-worlds-first-ai-robot-eve-8626164/
- https://ai.googleblog.com/2023/03/palm-e-embodied-multimodal-language.html
- https://www.prnewswire.com/news-releases/humanoid-robot-market-worth-17-3-billion-by-2027--exclusive-report-by-marketsandmarkets-301520783.html
- https://www.theguardian.com/technology/2022/sep/30/tesla-optimus-humanoid-robot-elon-musk-ai-day
- https://synchron.com/
- https://neuralink.com/
- https://blogs.scientificamerican.com/observations/dont-fear-the-terminator/
- https://indiaai.gov.in/article/andrew-ng-and-yann-lecun-on-why-ai-research-shouldn-t-be-paused
- https://www.gatesnotes.com/The-Age-of-AI-Has-Begun
- https://www.deepmind.com/blog/an-early-warning-system-for-novel-ai-risks
- https://openai.com/blog/governance-of-superintelligence
- https://tech.hindustantimes.com/tech/news/chatgpt-openai-ceo-sam-altman-ai-watch-dog-71686068947296.html
- Superintelligence: Paths, Dangers, Strategies - Nick Bostrom

Image References

- https://commons.wikimedia.org/wiki/File:The_Prime_Minister_meets_with_AI_developers.jpg - Image used in societal impact section
- https://pixabay.com/it/illustrations/cervello-mente-umano-umanoide-7412662/
- https://pexels.com - Pavel Danilyuk
- https://www.pexels.com/@pixabay/
- https://commons.wikimedia.org/wiki/File:Hand_off_(52535927777).jpg
- https://futureoflife.org/open-letter/pause-giant-ai-experiments/
- Midjourney

Cities Hitched to Informatics ? : Urban Analytics

- https://arxiv.org/abs/1706.05535
- https://arxiv.org/abs/2111.07489
- https://arxiv.org/abs/2012.14349

Image References

- https://indianexpress.com/article/cities/bangalore/bengaluru-namma-metro-world-resources-institute-technical-advisors-7482932/
- flickr.com - David Orban (Picnic)
- https://www.wricitiesindia.org/content/data-led-urban-planning

References

- https://unsplash.com/photos/PM_VwL2ypes
- https://unsplash.com/photos/csC6CfiYKF0
- https://medium.com/geekculture/the-energy-transfer-problem-charging-electric-vehicles-ed-662c2a3cd6
- https://m.facebook.com/sidewalklabs/events/
- https://www.dezeen.com/2020/10/20/delve-sidewalk-labs-machine-learning-tool-cities/
- https://uk.urbanest.com/journal/best-parks-east-london/

Breaking Boundaries with Synthetic Data

- https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data/

Image References

- https://commons.wikimedia.org/wiki/File:Float_Glass_Unloading.jpg
- https://www.pexels.com/photo/warehouse-with-concrete-floors-4483610/
- https://www.researchgate.net/publication/351752306_Review_on_Generative_Adversarial_Networks_Focusing_on_Computer_Vision_and_Its_Applications
- https://pexels.com : Matthias Groeneveld
- https://www.wallpaperflare.com/cube-mesh-abstract-big-data-and-data-mining-concept-with-copyspace-wallpaper-aecoc
- https://elise-deux.medium.com/everything-that-happened-in-the-synthetic-data-space-in-2022-c5d6cb5aaf06

Privacy Preserving Techniques in ML

- https://arxiv.org/pdf/2004.12108.pdf
- https://arxiv.org/pdf/2008.03686.pdf
- https://arxiv.org/abs/2304.03442

Image References

- www.freepik.com
- www.pixabay.com

The Creative Renaissance: How LLMs are shaping generative AI

- https://www.nvidia.com/en-us/glossary/data-science/generative-ai/
- https://finalboss.io/ai-in-music/
- https://docs.midjourney.com/
- https://www.google.com/search?q=musiclm+paper&sourceid=chrome&ie=UTF-8
- Papers:
- https://arxiv.org/pdf/2108.07258.pdf
- https://arxiv.org/pdf/2305.10403.pdf
- https://arxiv.org/pdf/2301.11325.pdf

References

Image References

- https://www.pexels.com/photo/laptop-office-working-internet-16094047/
- https://www.trustedreviews.com/explainer/what-is-google-palm-2-4325830
- https://playgroundai.com/post/cllc1tlez0j5cs601sos1ys3v
- https://playgroundai.com/post/cllmxk4t606hks6015zd1dzti
- https://commons.wikimedia.org/wiki/
- https://www.flickr.com - CErixsson - concert

Algorithms to Allies: Shaping the Future of Policing

- https://www.criminallegalnews.org/news/2020/aug/15/problems-predictive-policing/
- https://techdetector.de/stories/predictive-policing
- https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained
- https://d3.harvard.edu/platform-rctom/submission/using-machine-learning-for-crime-prediction/
- https://blogs.lse.ac.uk/humanrights/2021/04/16/predictive-policing-in-india-deterring-crime-or-discriminating-minorities/
- https://vce.usc.edu/volume-5-issue-3/pitfalls-of-predictive-policing-an-ethical-analysis/
- https://vidhilegalpolicy.in/blog/indias-tryst-with-predictive-policing/
- https://www.predpol.com/technology/

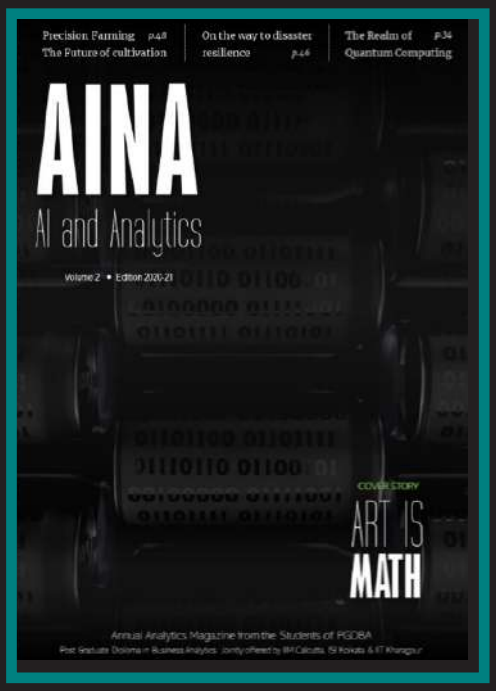
Image References

- https://commons.wikimedia.org/wiki/File:Seal_of_the_Los_Angeles_Police_Department.png
- https://www.theguardian.com/us-news/2021/nov/07/lapd-predictive-policing-surveillance-reform
- https://commons.wikimedia.org/wiki/File:High_Court_of_Telangana_in_Hyderabad04.jpg
- https://commons.wikimedia.org/wiki/File:Hyderabad_city_police_logo.jpg
- https://www.pexels.com/photo/judges-desk-with-gavel-and-scales-5669619/
- https://flickr.com - Michael Coghlan - South Australian Supreme Court

Transformers

- https://github.com/Significant-Gravitas/Auto-GPT

PREVIOUS EDITIONS



To read AINA Vol 1 scan the QR code above or go to https://online.anyflip.com/lgfta/gxxz/mobile/index.html

To read AINA Vol 2 scan the QR code above or go to https://online.anyflip.com/ex-okr/ehwf/mobile/index.html

To read AINA Vol 3 scan the QR code above or go to https://anyflip.com/lgfta/wqew