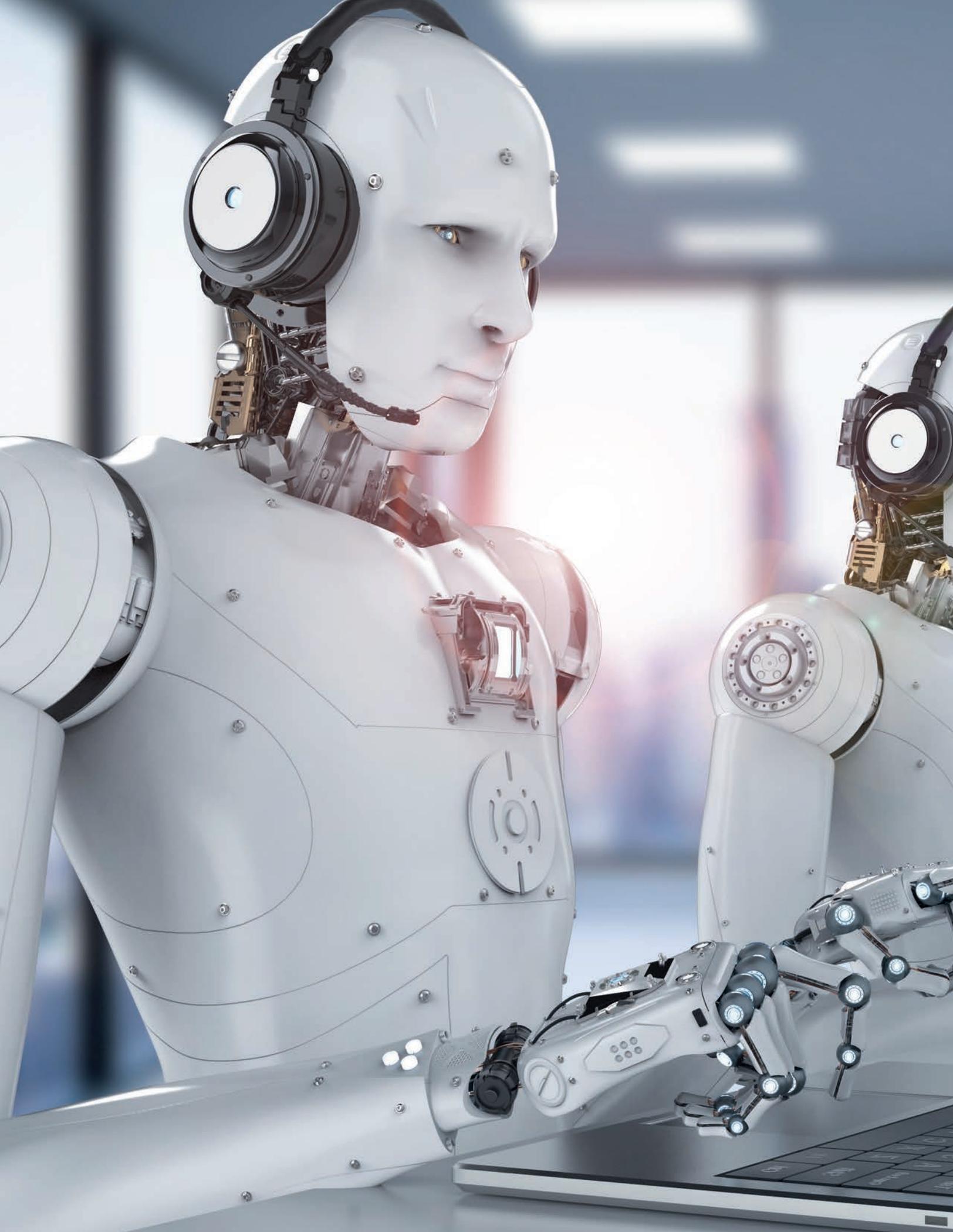
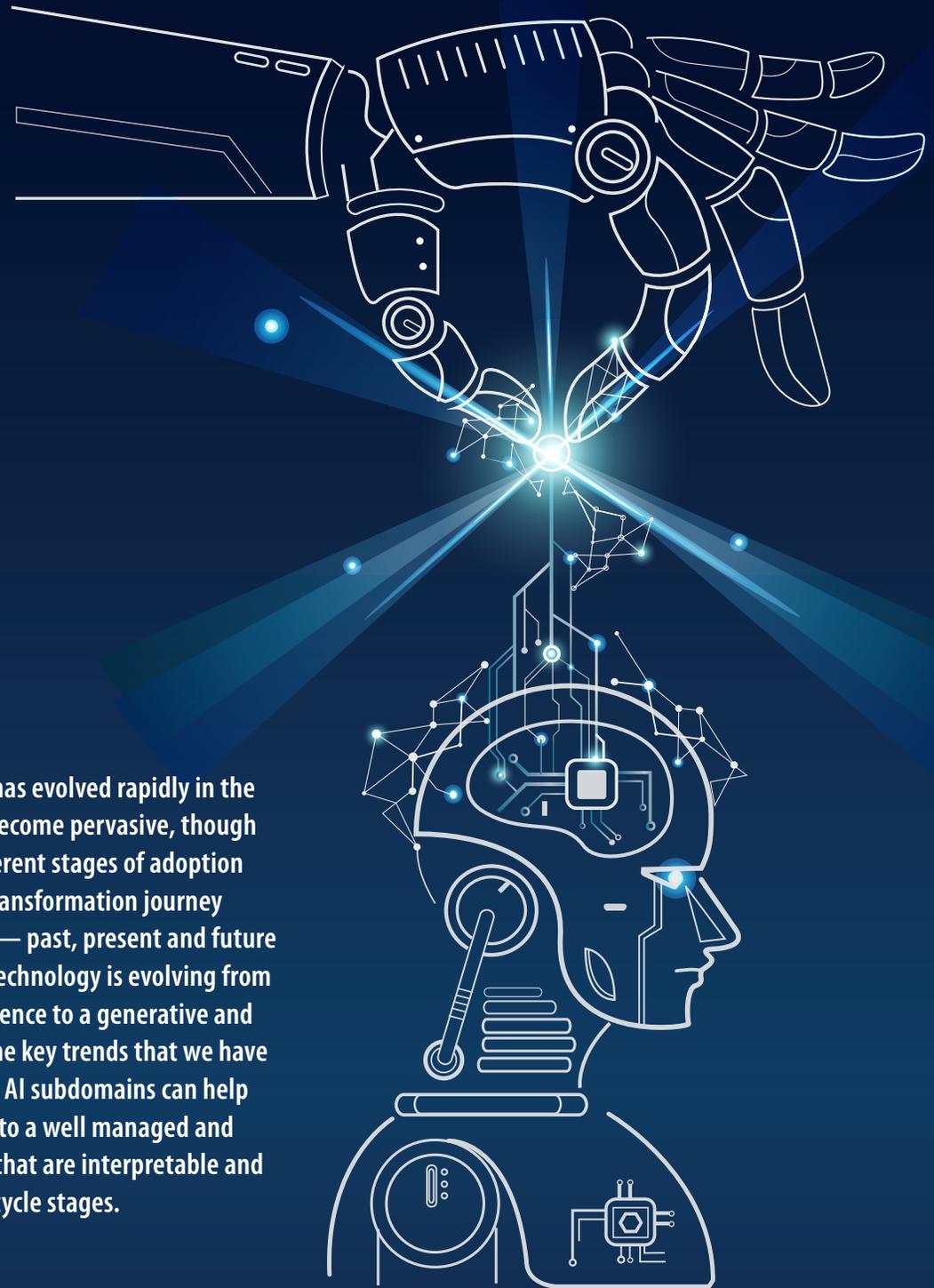


AI-FIRST TO
BE A LIVE
ENTERPRISE

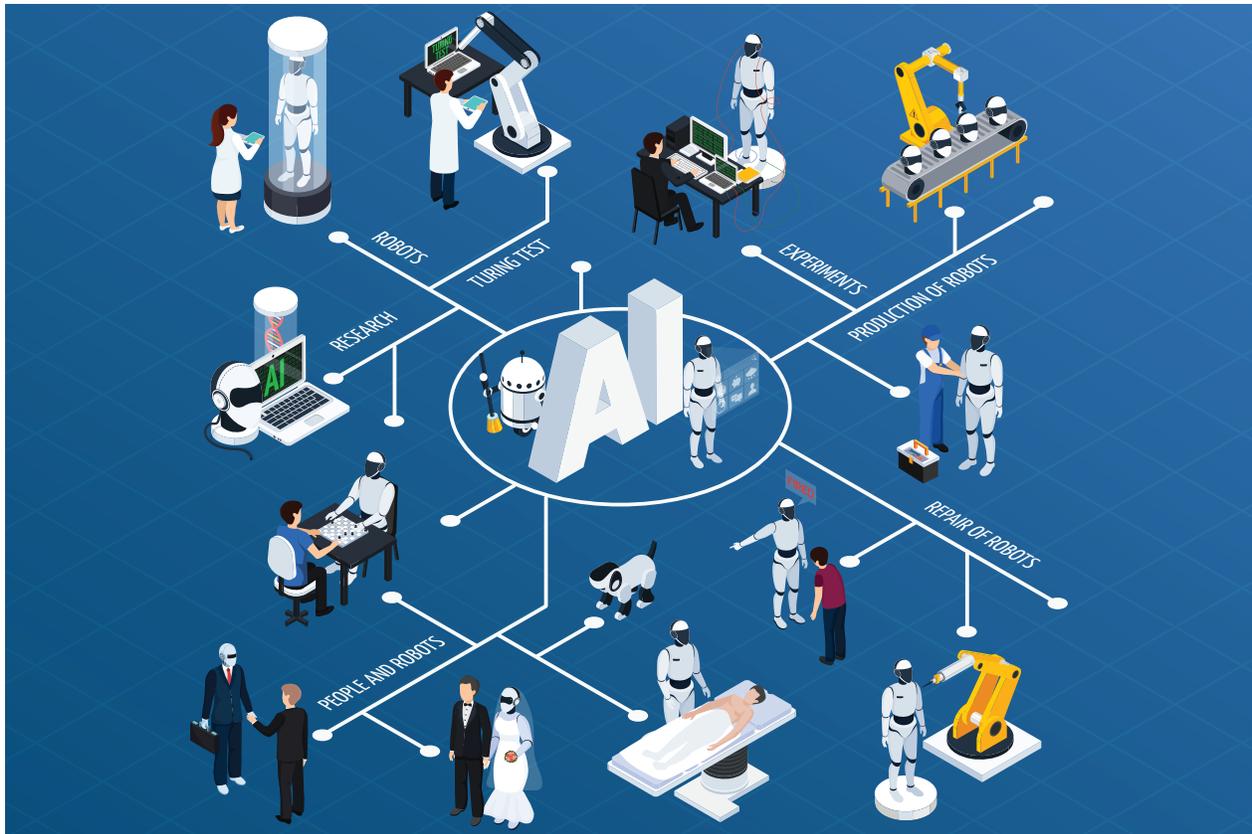


Contents

Evolve to an AI-first live enterprise	5
AI algorithms and architectures	8
Computer vision	10
Speech	12
Natural language processing	15
AI on the edge	18
AI life cycle tools	20
AI governance	21
Adopt H3 to move to a live enterprise	23
References	25
Advisory council and contributors	26



Artificial intelligence has evolved rapidly in the last few years. It has become pervasive, though enterprises are at different stages of adoption and AI usage. The AI transformation journey across three horizons — past, present and future — explains how the technology is evolving from an augmented intelligence to a generative and explainable system. The key trends that we have discussed across seven AI subdomains can help enterprises transform to a well managed and -governed AI systems that are interpretable and explainable at all life cycle stages.



The COVID-19 pandemic has brought real urgency to the adoption of touchless technologies like facial recognition, speech biometrics or flagging social distancing non-compliance. There has never been more need to fast-track the drug discovery and trial process than is demanded today. AI clustering algorithms can potentially help group similar profiles of virus and DNA vaccine pairs. From millions of data points in databases, they can help locate virus DNA structures closer to a target such as COVID-19. Creating simulation pathways of DNA trial research and corresponding treatments can be used as a base to accelerate COVID-19 vaccine discovery and treatment paths.

The digital AI-first transformation enables organizations to create competitive advantages and develop brand-new products, services and business models. With the ability to sense changing employee, partner and customer dynamics, the enterprise of the future will use AI intelligently and at scale.

Evolve to an AI-first live enterprise

As enterprises undergo this AI transformation, they are moving across three horizons.

Horizon 1 (H1) systems are characterized by augmenting fragmented intelligence into existing systems with capabilities such as customer recommendation and fault prediction, usually implemented using classical AI algorithms like Naive Bayes, Support Vector Machines and Random Forest.

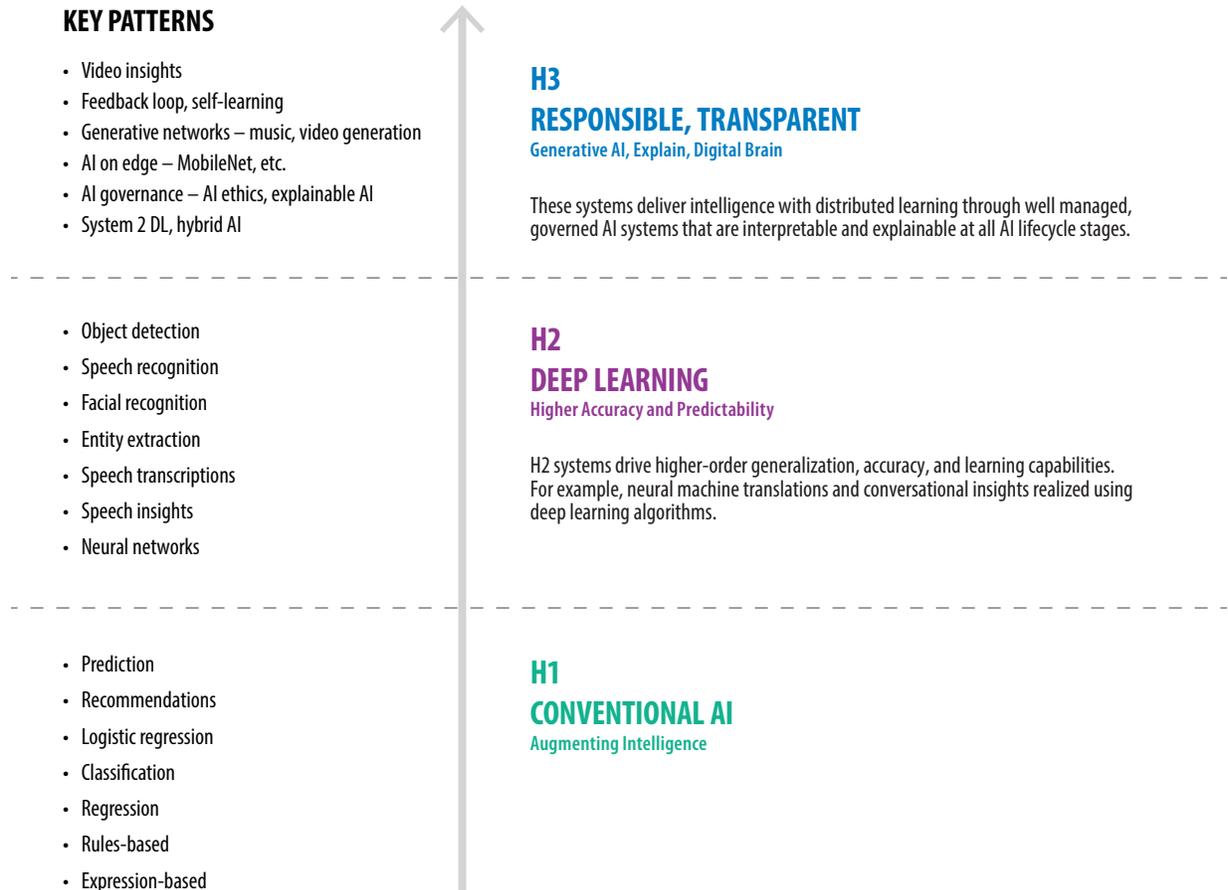
H2 are complex systems that need higher-order generalization, accuracy, and learning capabilities, for example neural machine translations and conversational insights realized using deep learning algorithms.

H3 systems drive toward semi-supervised to unsupervised, transparent, multi-task learning

systems. Semi-supervised machine intelligence-based examples can have a generative capability for textual, audio and video content; video-based insights generation, such as activity recognition or video summarization. Leveraging proliferation of sensors,

cameras and devices, they deliver rich intelligence at edge with distributed or federated learning. More importantly, it aspires to achieve all this through well managed, governed AI systems that are interpretable and explainable at all life cycle stages.

Figure 1. Adapting to market dynamics: the three horizons



Source: Infosys

Moving to H3 will need enterprises to work across one or more of the following AI subdomains

1. AI algorithms and architectures
2. Computer vision
3. Speech
4. Natural language processing
5. AI on the edge
6. AI life cycle tools
7. AI governance

Figure 2. Key trends across AI subdomains



Source: Infosys

In this paper, we explore key trends under each of the subdomains that can help transform an enterprise from a fragmented, ad hoc intelligent entity to a creative, efficient, responsible intelligence-driven ecosystem.

AI ALGORITHMS AND ARCHITECTURES



Adoption of deep learning and transfer learning architectures will drive accuracy, performance and speed

Trend 1: Improve generalization and accuracy with deep neural network architectures¹

Adoption of deep learning-based solutions to solve enterprise-class problems is driven by some key factors, such as availability of graphics processing unit computing (GPU), availability of large labeled data, and fast-paced innovations in new deep learning algorithms. They promise higher accuracy and better

generalization characteristics as compared to classical algorithms such as SVM, Naive Bayes and Random Forest. However, the need for a large set of labeled data and the cost of GPU computing are still two challenges to the mainstream adoption of deep learning. Transfer learning-based models, have made huge headway in overcoming the limitations of lack of sufficient availability of labeled data and GPU. Addressing certain complex problems in computer vision, NLP and speech domains has become feasible with evolving architecture such as Transformers.

Infosys partnered with a large technology company to transform its existing system that was moderating user uploaded content based on certain pre-configured historical rules and policies to an AI-based moderation. The AI model was trained with a very limited set of label data for supervised transfer learning-based deep neural net architecture for vision and text to identify, classify and isolate any toxic content arriving from user-uploaded forms.

As part of a prestigious global tennis tournament, Infosys trained a transfer learning-based computer vision model using a limited set of brand logo images, to establish the amount of time and time frames a particular brand was visible. Similarly, for compiling key moments, player action identification such as waving hands to the crowd was done with a similar approach.

There are several examples illustrated across the following sections where Infosys deployed transfer learning-based techniques for speech, vision and text to overcome labeled data deficit challenges during supervised model training. In some cases, Infosys also used conventional AI algorithms such as logistic regression, SVM to drive early results and then used those results to train a deep learning-based AI model for improved generalization and accuracy.

Trend 2: Transition from System 1 deep learning to System 2 deep learning²

The current state of deep learning-based AI is referred as System 1 deep learning, and it can be best illustrated with an example of a person driving a car in a known vicinity while talking on the phone or with a passenger, and is able to automatically drive through, without consciously focusing on driving. However, the same person driving through an unknown vicinity will need a lot more focus and will need to use various logical reasoning and connections to reach the destination. These types of problems, which need a combination of reasoning and a sense of on-the-fly decision-making, still can't be solved with current AI discipline maturity and are considered System 2 deep learning.

System 1 deep learning's current state is due to certain current limitations of deep learning's generalization capabilities, where these algorithms

- are not able to correctly work on (detect) unseen data patterns;
- need to have balanced distribution of data in training and testing sets;
- lack the ability to do continuous learning based on changes in environments in real time, similar to active agents;
- lack the logical and reasoning capability to combine high-level semantic concepts, and
- are unable to deal with out-of-distribution (noise) data.

These are some of the reasons for the current state of AI's inability to deal with the System 2 deep learning state.

System 2 deep learning is where some of these challenges are being worked upon by leveraging techniques like attention-based architectures and models, multitask learning, incorporating principles of consciousness, and meta learning with an emphasis on unsupervised, zero-shot learning techniques.

COMPUTER VISION

Graduation from mere object identification to deep learning-based video insights

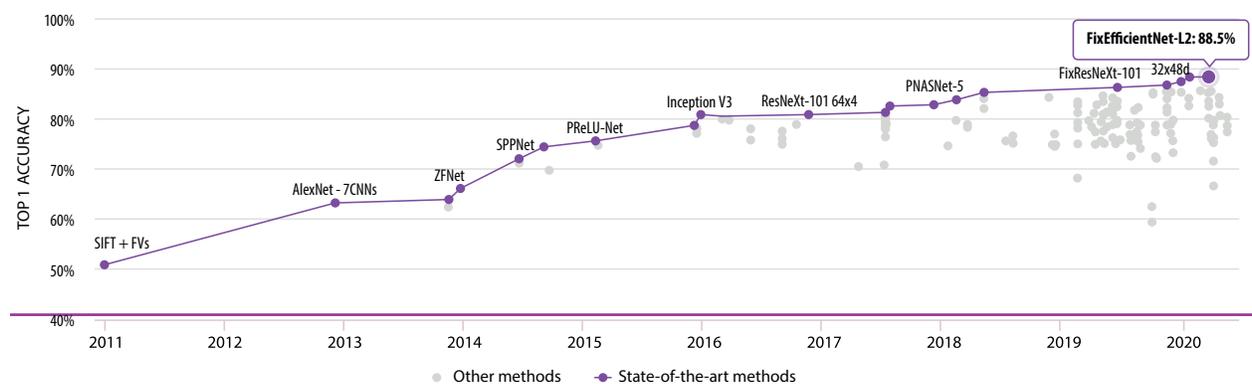
Leverage of Computer vision (CV) implementations within the enterprise to solve image- and video-based insight problems has improved significantly with the availability of state-of-the-art pretrained open-source models and the evolution of neural network architectures. Since 2010, with the global ImageNet Large Scale Visual Recognition Challenge, the image classification error rate has decreased from >25% to <2% and has surpassed the human level of accuracy. This is possible because of the continuous advent and evolution of new convolutional neural

network architectures, such as AlexNet, VGG, GoogleNet, ResNet and ResNeXt-50 to the most current FixEfficientNet-L2, with an error rate of less than 2% in 2020. These competitions have consistently proved that accuracy and performance are not solely dependent on the volume of training data; however, innovative crafting of neural net architectures has played a differentiating role³

The evolution of various state-of-the-art neural net architectures can be best depicted using this graphic.

Figure 3. Evolution of various state-of-the-art neural network architecture(s) from ImageNet competitions⁴

Image classification on ImageNet



Source: Papers With Code

Trend 3: Image segmentation, classification and attribute extraction

Object detection, segmentation and classification are the building blocks to address several complex computer vision challenges. Object detection helps to identify an object in the image, forms a rectangular boundary and creates a bounding box to narrow down the object. Image segmentation then identifies the object with all the curves, lines and exact shape

of the object. This helps in more granular and finer identification as against simple object detection. Object classification helps classify a particular object into a class or subclass. For example, classifying a vehicle into a car or an airplane and further sub classifying the brand into Audi, BMW, etc.

Object detection, segmentation and classification help in the healthcare sector by identifying and narrowing tumor regions and further classifying the tumor as malignant or not.

These also help to establish various insights from images by classifying them to categories, segmenting them for specific information, or extracting any image attributes.

Infosys partnered with a large global energy company to identify faulty cables based on the picture sent from the site of failure, which allowed them to send the right engineer to the site to fix the cable. This helped the company in saving costs from sending engineer to fix wrongly reported cable problem.

Infosys worked with a large global retailer to extract and classify information from digitally scanned product art (SmartArt) so that the information was extracted correctly and could be further classified as contents, ingredients, instructions, etc. to make the information available on multiple channels for regulatory and compliance purposes.

Trend 4: Video insights

There are several interesting possibilities emerging from applying AI to videos, such as generating video captions, video highlights, content moderation, span of brand coverage, surveillance, and people or object tracking.

As part of a prestigious global tennis tournament, using various CV-based algorithms, Infosys extracted various game insights to create highlights from various events, such as seeing players waving to the crowd, extracting the score from the video feed, recognizing players and determining the length of the time a particular advertisement or brand was featured in a video.

Similarly, for a large railroad company in the U.S., various assets spread across geography were identified and counted from a streaming video feed obtained using a train-mounted camera.

Some of the new problems we are working on with clients in the CV space include handwriting recognition in the context of know-your-customer forms that are written manually and need to be digitized, activity and pose recognition in a video,

video synthesis, video summarization, and image captioning by leveraging state-of-the-art AI models and techniques such as 3D object detection, generative networks and single-shot learning.

SPEECH



Context-specific models will drive enterprise adoption of speech-based experiences

Significant amounts of intelligence and insight are buried deep in conversations, whether they are conversations between a customer and a contact center executive or conversations between a customer and a trading desk executive.

Huge amounts of conversational data are generated in real time in enterprises that can be tapped into in order to derive intelligence to improve the quality of conversations, whether pertaining to resolving customer issues for higher satisfaction, identifying and tapping into product feedback opportunities, reducing costs by identifying frequently asked questions and moving them to self-service channels, or using them merely to train and engage the staff better.

Opportunities in enterprise are immense, as voice-based, unstructured conversations are becoming the next big source of intelligence after emails and documents. However, there is one problem in deriving intelligence from these conversations — they need to

be cleanly transcribed, and this is the first step on the journey of deriving intelligence.

To achieve clean transcriptions, there are several technical challenges, such as disparate speaker languages (e.g., English, Chinese, French); conversation-centric vocabulary; varying accents; ambient noise; and also different channels, such as mono and stereo, used for recording conversations.

Over many years, large players like Microsoft, Google and IBM have gathered a huge corpus of voice data and created proprietary speech-based models that can transcribe these conversations at fairly high quality. The only challenge with them is that voice data needs to be sent over to the cloud, and at times, many customers are not keen to send their data because of confidentiality and privacy concerns. Also, if speech models need any domain-specific custom adaptations, these cloud players have problems with training limitations.

On the other side, there are several open-source engines, including Kaldi, CMU Sphinx, Mozilla DeepSpeech and Facebook's wav2Letter, that provide the open-source rich models with the ability to transcribe. Their advantages are, they provide different granularity of the ability to train custom speech models. This in turn can help drive higher accuracy of transcriptions for domain-sensitive models. The control, granularity of training data, effort required and accuracy outcomes are different for different open-source players. This evolution is driving our next AI enterprise adoption trend.

Trend 5: Adoption of neural machine translation- and transcription-based systems to mine conversational insights

Historically, translation systems have been implemented using Statistical Machine translations primarily using count-based models. They were best

suited for short sentences with standard nouns and phrases, importantly they are lightweight models. Neural Machine Translation and Transcription based systems have brought in significant improvement in accuracy and speed. Improvements are due to usage of deep learning, multi-head, self-attention mechanisms with encoder-decoder transformer architecture style on a pre-trained transfer learned corpus. Their model size is usually large with a parameter size of millions to billions and needs more than one GPU. They also make zero-shot learning possible. For example, in the absence of underlying language data for Portuguese translation, a translation from Spanish to Portuguese is achieved by translating it first from Spanish to English and then English to Portuguese.

For a large railroad company in the U.S., Infosys assisted in transcribing call center conversations using speech-based custom models to identify which product lines had maximum issues being reported, which agents were driving customer satisfaction versus which ones needed training, and the correlation of the rise or the drop in calls with events such as line failures, new product launches and faults reported.

Infosys partnered with a large global airplane manufacturer to transcribe the conversations between pilots and ground staff. The conversations were studded with cockpit noise, strong regional accents, different languages and heavy ambient noise. We successfully custom-trained for the variation in accents to deliver high transcription accuracy, ran language insights to infer causes of delay in flight landing and accidents in the air, and also used transcriptions and insights to improve ground staff and pilot training.

For a large retailer in the U.S., Infosys transcribed call center conversations to derive information about intentions, conversation sentiments and key topics, and we also clustered subject-specific intentions, such as all order delivery-related, payment-related and product return-related intentions in order to understand and improve gaps in supply chain operations.

Trend 6: Speech biometrics

Speaker-based authentication and verification is another key trend that is getting adopted as an augmented biometric method in addition to those already deployed by enterprises, such as using thumbprint or facial recognition. With the COVID-19 situation, this has gained more relevance.

Experiences that are now de facto with smartphone experiences — where Android and iOS operating systems are able to provide hooks to capture user voices, train and then use them for user authentication, search, query and other functions — are slowly graduating into the enterprise too.

Speaker verification and identification are used in several ways, such as identifying the caller and then greeting the person by name to make it highly personalized or pulling back-end data to provide contextual recommendations and suggestions without exclusively asking for the name; these have a significant influence on the experience of the user.

In this context, among others, we worked with a large global financial institution to develop speaker authentication in a contact center.



NATURAL LANGUAGE PROCESSING



Shift from extraction of isolated entities to abstractive reasoning

Enterprises have an unstructured data deluge of documents, images, emails, blogs, voice conversations and video data, and it provides a huge opportunity to mine intelligence. On the other hand, there is an important subdiscipline in AI as NLP is undergoing significant innovations to improve machines' ability to understand (NLU), generate (NLG) and process and derive insights (NLP). Similar to vision and speech, transfer learning-based training of text is playing a significant role in getting the NLP-based models running quickly.

The early use of NLP (H1) was centered primarily on extracting and representing information as a bag of words with hot encoding-based sparse vectors, where emphasis was given to individual words rather than the organization of words in sentences, which resulted in loss of meaning. Also, it extracted named entities, such as organization name, person name, location, date and time from the text with nearly no ability to train for custom entities such as currency symbols.

NLP evolution that we refer to as H2 moved further, with the ability to train for custom-named entities such as currency symbols. Adoption of deep learning-based word vector models such as GloVe and Word2Vec helped establish word similarity and synonyms without needing to train for every word in the dictionary. However, they soon introduced various limitations, such as the ability to deal with out-of-vocabulary words, misspelled words, sentences, paragraphs, and language context-related challenges such as entity coreference resolutions. Due to a lack of contextual information in embeddings and needing to rely on distance between the vectors, this ended up introducing word distance-based biases.

Transfer learning-based text training is getting increasingly vital in speeding up NLP-based models

Long-term short-term memory architectures (LSTMs) overcame these limitations and were able to capture the long-range dependency through memory gates (cells) and deal with coreference resolution problems; however, they took a very long time to process because of a lack of parallel processing capability.

Attention networks' ability to focus on a specific part of an input sentence helped reduce the training time with parallel processing and was a significant improvement over LSTMs.

Transformers with encoding and decoding architecture are special types of attention networks. They can be trained in parallel, deal with longer sentences, and are a lot faster to train and predict. They are the architectural building blocks behind all the modern state-of-the-art results provided by various models — Google's BERT, XLNet, Open AI's GPT-2, Facebook's RoBERTa, Baidu's ERNIE 2.0 and Google's T5. Since then, there have been several variations of BERT-based architectures that have come up, an important one being to process multimodal video data in a self-supervised way (Video BERT) and

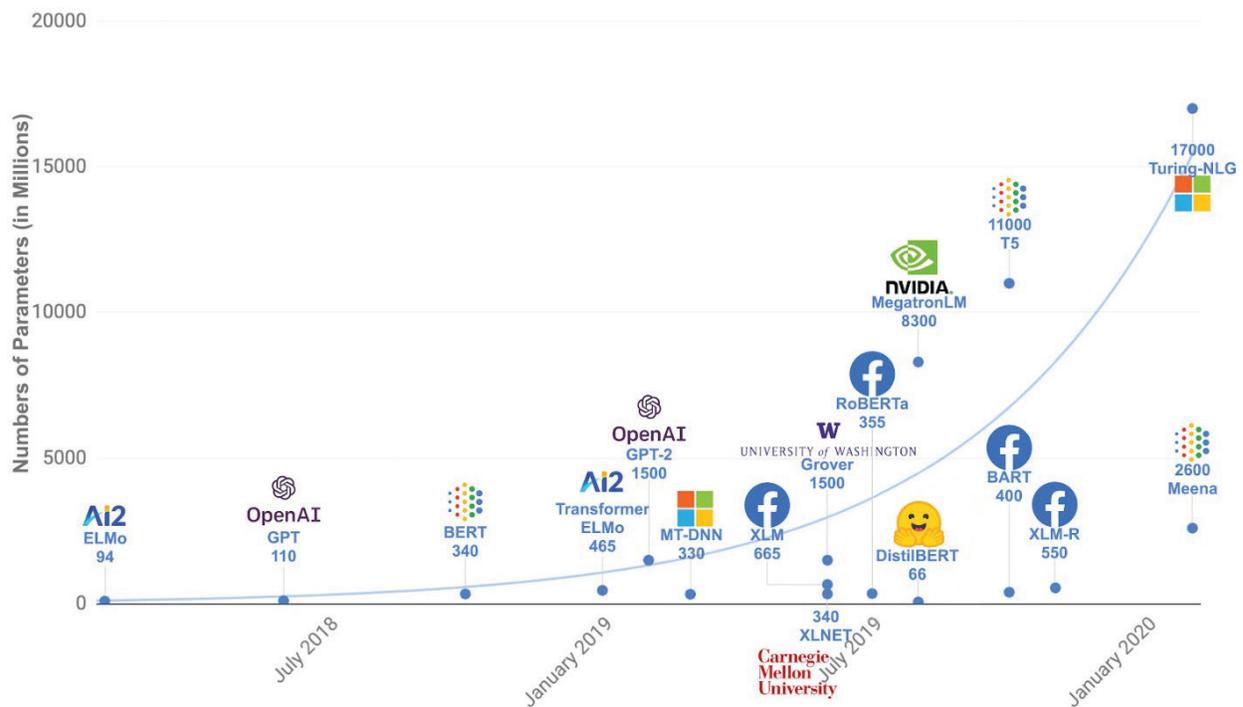
one jointly learning language and vision (ViLBERT).

ERNIE 2.0 supports multitask learning and has proved to have better accuracy than BERT and XNLET on 16 standard language tasks on general language understanding evaluation benchmarks and certain other Chinese tasks. SuperGLUE, the advanced version of GLUE with more complex tasks, is the new competition ground for NLP tasks, and presently, Google's Text To Text Transfer Transformer (T5) is the leader, with the highest accuracy for all 10 tasks.⁵

Microsoft's Turing-natural language generation (Turing-NLG) is by far the latest and largest model, with 17 billion parameters. It outperforms the state-of-the-art models on a variety of language modeling benchmarks and also excels in numerous practical tasks, including summarization and question answering, and it has the ability to also do zero-shot learning.

The graphic below depicts the evolution of various state-of-the-art NLP models and their corresponding sizes.

Figure 4. The evolution of various state-of-the-art NLP models⁶



Source: TensorFlow Blog with reference from DistilBERT paper

H3-based implementations in NLP are starting to take shape with usage of previously discussed state-of-the-art transformer-based architectures and models, leveraging contextual and crosslingual word embeddings, and using multitask learning to solve various language problems.

NLP has play in several areas — autoclassification, clustering of documents, extracting key entities and paragraphs, or doing sentiment analysis of text. The next key trend is driving AI adoption in enterprise.

Trend 7: Derive content intelligence from forms extraction, document attributes and paragraphs

Enterprises have information embedded in various types of documents and in the form of digital or handwritten content. These include research study documents, Know Your Customer forms, payslips and invoices. Extracting key information points and systematically digitizing this information are key problems and the driving pattern across various industries.

Infosys worked with a large global pharmaceutical company and used NLP techniques to extract various product characteristics — such as the chemical composition of drugs, posology, severity and comorbidity — from clinical research documents.

Infosys partnered with a large global seed manufacturer to extract various information data points from intellectual property documents related to studies and details of various experiments that were spread across geographies in different shared locations, languages and versions.

Infosys assisted a large bank with a solution that digitized information received from various vendors in the form of invoices that were in different formats and file types.

At Infosys, several data points get captured as part of client contracts. We are extracting sensitive financial information, such as contract name, value, start date, end date and other sensitive clauses, such as liability and indemnity, from contract documents using AI-based techniques.

AI ON THE EDGE



Elevation from on-device intelligence to federated intelligence

Trend 8: Address latency, point-specific contextual learning with edge-based intelligence

Smart Reply, auto suggestions for grammar, sentence completion while typing on a phone, voice recognition, voice assistants, facial biometrics to unlock a phone or an autonomous vehicle navigation system, robotics, augmented reality applications — all of them use local, natively deployed AI models to improve the response time to user actions. Imagine,

in the absence of a local AI model, the inference or prediction would have to be based on a remote server; the experience would be just completely opposite. AI plays a key role in providing improved experience to the user by leveraging edge-based AI.

Edge-based AI plays a quintessential role in remote locations, where network connectivity may not be continuous, response times should be in fractions of seconds and network latency cannot be afforded, and hypercontextualization is required with user-specific data in the given environment.

Edge-based AI is feasible because of a significant improvement in edge processing specialized embedded chip hardware and software such as Google tensor processing unit, field programmable gate arrays and GPU.

At the edge, typically two things happen, one being inference or prediction and the second being training or learning. For the inference or prediction to happen, a lightweight model is available to predict. The model with training capability can use local context-based learning and at the appropriate time can synchronize with the central model. The synchronization can be done by just sharing the model parameters, weights, features, etc. without needing to share the actual data, thus managing data privacy. Once the central model builds itself with several such updates from different remote edge-based AI models, it can update its training and share the updated model footprint with all the edge-based devices or clients, thus ensuring

everybody gets the benefit of the central learning capacity. This process of distributed learning is called federated learning, and essentially it is employed as a strategy where sharing data has challenges of data privacy, shareability, network transport limitations, etc. but at the same time needs to leverage the benefits of abstracted learning available through the central capacity.⁷

TensorFlow Lite provides the complete toolkit to convert TensorFlow models to TensorFlow Lite, which can run on edge devices. They also are made compatible to gain the benefits of central processing unit and GPU acceleration devices. MobileNet models adapt several state-of-the-art CNN models to device models by sizing network architecture patterns such as depthwise separable convolutions, hyperparameter optimization for width multipliers, and resolution multipliers with the corresponding trade-offs in accuracy and latency.⁸

Infosys partnered with a large European car manufacturer to use edge computing to identify and predict failures of spindle machines in a brownfield environment through an “internet of things” gateway. This helped in setting up a cost-effective solution for handling the large data feeds coming from the spindle machine.

A large global mining company used wearable devices for safety monitoring of moving miners’ source data to the cloud for processing on a near-real-time basis.

For a large global manufacturer, optimization of the engine maintenance, repair and overhaul shop floor is driven through edge computing, which provides the availability and predictability of the machines for scheduling operations.

AI LIFE CYCLE TOOLS

Shift from fragmented to integrated, managed and monitored pipeline tools

Trend 9: Integrated AI life cycle tools to drive enterprisewide standardization

The AI life cycle involves various stages, from data collection, data analysis, feature engineering and algorithm selection to model building, tuning, testing, deployment, management, monitoring and feedback loops for continuous improvement. As such, the software development tools and processes are fairly standardized with DevOps. However, with the breadth and depth of AI disciplines, frameworks and languages, specialized capability is required and specialized tools need to come in, to manage every stage of AI project development. The software space is also fairly fragmented, with tools from large companies to small startup players.

Based on our interactions with clients, we are starting to see adoption of end-to-end AI life cycle development tools including H2O.ai, Kubeflow, and MLflow in enterprises; however, there is a long way to go, as standardization of these tools and pipelines is still a work in progress.

Trend 10: Model sharing and reusability through model exchanges

Creating an AI model from scratch needs a huge amount of effort and investment for collecting datasets, labeling data, choosing algorithms, defining

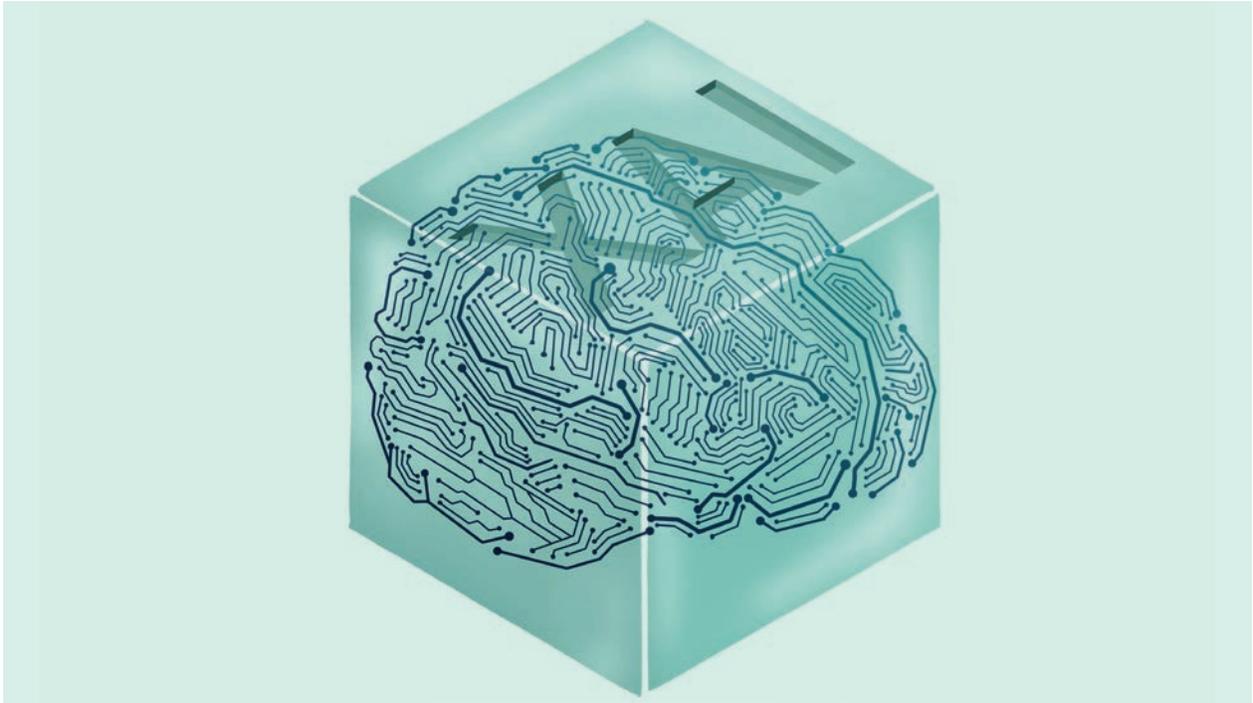
network architecture, establishing hyperparameters, etc. Apart from this choice of language, frameworks and libraries along with client preferences, etc. differ from one problem space to another.

With these challenges, it is important to have a way to reuse the effort invested across the community by sharing models and ensuring model compatibility and portability across environments.

This brings up the need for a minimum of two mechanisms. One is a place where models can be shared for global usage. The global context can be truly global, across enterprises, or within enterprises. Currently, most of the models that exist are basic in nature, such as in vision-object detection and activity recognition. However, there is a need to share the models and datasets that are specific to domain problems for health care, finance, insurance, energy, etc. The second type of mechanism is where models built using Python, TensorFlow and Cuda versions on specific types of GPUs or CPUs need to be compatible and portable to and from PyTorch or other frameworks, libraries and environments.

Open neural network exchange (ONNX)⁹ is one such leading open-source model exchange that hosts various pretrained models using ONNX Model Zoo. Similarly, TensorFlow Hub and Model Zoo provide various datasets and models created by the TensorFlow community.¹⁰

AI GOVERNANCE



Shift from black box to interpretable systems

Trend 11: Adherence to AI ethics as a underlying principle to build AI systems

With the adoption of AI systems increasing in critical decision-making systems, the outcomes rendered by these systems become critical. In the recent past, there have been examples where the outcomes were wrong and impacted important human issues, some of the examples being an AI hiring algorithm found to be biased against specific races and gender. Prison sentences had twice as high a false positive rate for black defendants as for white defendants. A car insurance company by default classified males under 25 years as reckless drivers.

AI started attracting a lot of negative press because of failing systems, the resulting lawsuits and other societal implications, to the point that today regulators, official bodies and general users are seeking more and more transparency of every decision made by AI-based systems. In the U.S., insurance companies need to explain their rates and coverage

decisions, while the EU introduced the right to explanation in the General Data Protection Regulation.

All the above scenarios call for tools and techniques to make AI systems more transparent and interpretable.

Following are some of the important principles required for operating an ethical AI system. To ensure trust and reliability in AI models, it is essential to adhere to some **key principles**:

- **Human involvement:** Though AI models are built to operate independently without human interference, human dependency is a necessity in some cases. For example, in fraud detection or cases where law enforcement is involved, we need some human supervision in the loop to check or review decisions made by AI models from time to time.
- **Bias detection:** An unbiased dataset is an important prerequisite for an AI model to make reliable and nondiscriminatory predictions. AI models are being used for credit scoring by banks, resume shortlisting and in some judicial

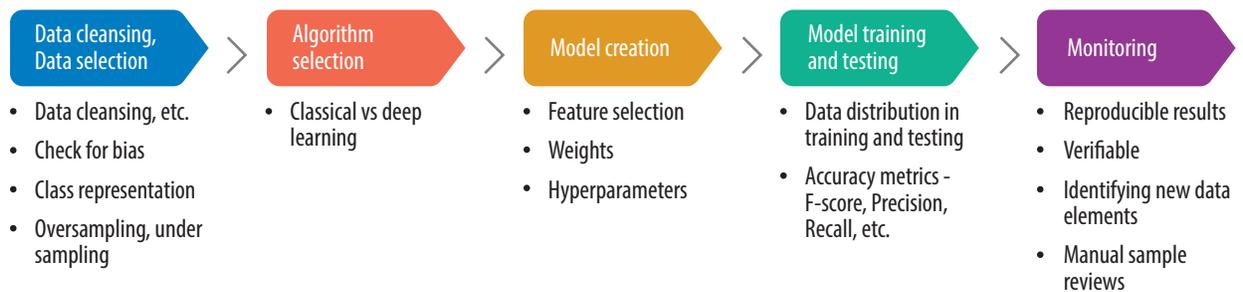
systems; however, it has been noticed that in some cases, the datasets had some inherent bias in them for color, age and sex.

- **Explainability:** Explainable AI comes into the picture when we talk about justifiable predictions and feature importance. Explainable AI helps in understanding how the model is thinking or which features of the given input it is emphasizing while making predictions.

- **Reproducibility:** The machine learning model should be consistent every time when giving predictions, and it should not go haywire when testing with new data.

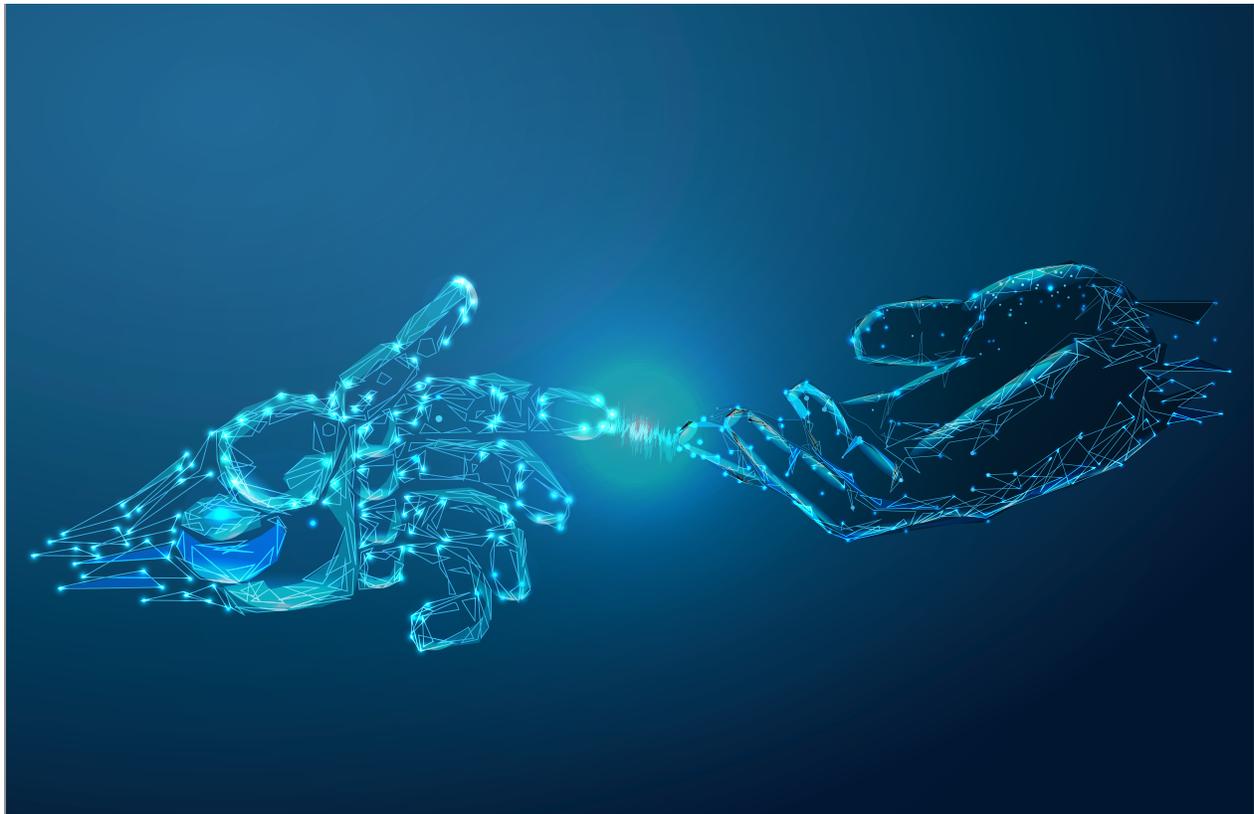
Many practitioners would mistake **explainable AI (XAI)** as being applied only at the output stage; however, the role of XAI is throughout the AI life cycle. The key stages where it has an important role are as follows.¹

Figure 5. XAI in the AI life cycle¹¹

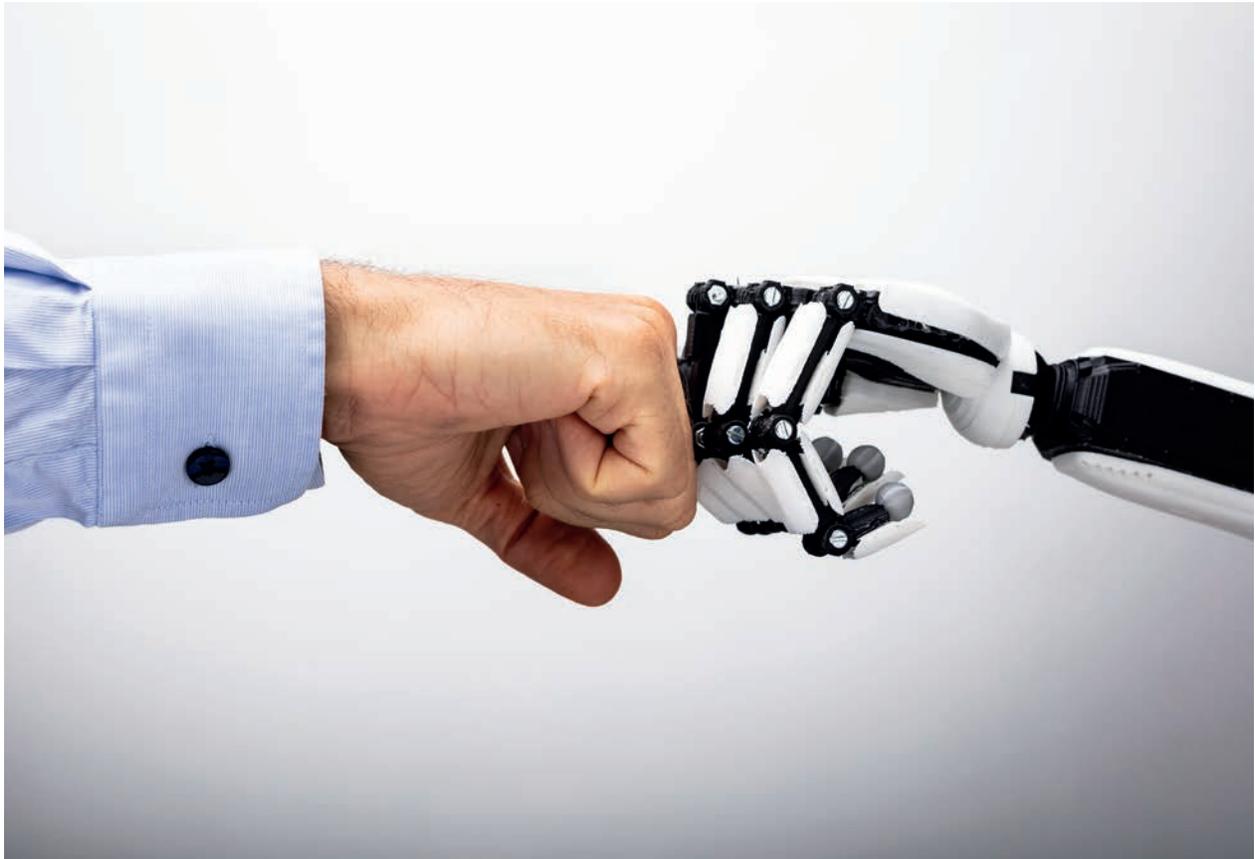


Source: Infosys

Thus, AI systems need consistent and continuous governance to make them understandable and resilient in various situations, and enterprises must ensure that as part of the AI adoption and maturity cycle.



Adopt H3 to move to a live enterprise



The current AI paradigm with neural networks at the heart is heavily dependent on the high volume of labeled training data, high compute and algorithmic innovations with their underlying complexities of describing the inferences. All these have made significant headway through transfer learning techniques; lightweight, highly tuned and optimized neural network architectures and models; and improved model explainability techniques. However, despite all of this, the neural network's ability to correctly respond to class of data that it hasn't seen during training gets severely constrained, has been an important area of research and is envisaged as a System 2 deep learning stage.

Though we have independently covered each domain and corresponding client implementation examples, there are many AI implementations where a fusion of one or more AI domains have been leveraged or should be utilized to deliver the client experience.

An example is in a content moderation implementation for a large US company, we used a live video stream to establish objectionable visual objects and also transcribed the stream with a neural language model to capture any objectionable sound profiles. Further, we processed this transcription through NLP to establish any objectionable words or phrases. All this was explainable and interpretable for every content type and was achieved through a seamless application programming interface-based AI microservices pipeline. This helped meet the AI goal of identifying any objectionable content being streamed live and also brought in a semi-automated way with a human in the loop.

The current state of AI still provides numerous opportunities for enterprises. Adoption of AI across domains and the journey toward H3 can help rapidly change the enterprise's response to the environment so it becomes a live enterprise.



References

Trend: Algorithms and architectures

- ¹ H3 trends in AI algorithms: The Infosys way
<https://www.infosys.com/services/incubating-emerging-technologies/offerings/documents/ai-algorithms.pdf>

Trend: System 1 to System 2 deep learning

- ² Yoshua Bengio video talk
<https://bdtechtalks.com/2019/12/23/yoshua-bengio-neurips-2019-deep-learning/>

Trend: Computer vision

- ³ ImageNet Large Scale Visualization Recognition Challenge (LSVRC)
<http://www.image-net.org/challenges/LSVRC/>
<https://www.kaggle.com/c/imagenet-object-localization-challenge>
<https://paperswithcode.com/sota/image-classification-on-imagenet>
<https://arxiv.org/pdf/2003.08237v4.pdf>
- ⁴ Figure: Computer vision architecture evolution
<https://paperswithcode.com/sota/image-classification-on-imagenet>

Trend: NLP

- ⁵ SuperGLUE benchmarks
<https://super.gluebenchmark.com/tasks>
Multi-Task Deep Learning Neural Network for Natural Language Understanding
<https://arxiv.org/pdf/1901.11504.pdf>
<https://github.com/namisan/mt-dnn>
T5 (Text To Text Transfer Transformer)
<https://arxiv.org/pdf/1910.10683.pdf>
ERNIE 2.0
<https://arxiv.org/abs/1907.12412>
- ⁶ Figure NLP Model evolution
<https://blog.tensorflow.org/2020/05/how-hugging-face-achieved-2x-performance-boost-question-answering.html>
<https://arxiv.org/pdf/1910.01108.pdf>

Trend: Edge computing

- ⁷ <https://arxiv.org/pdf/1907.08349.pdf>
- ⁸ MobileNet
<https://arxiv.org/pdf/1704.04861.pdf>

Trend: AI life cycle tools

- ⁹ Model exchanges
<https://onnx.ai/>
<https://github.com/onnx/models>
- ¹⁰ TensorFlow Hubs
<https://www.tensorflow.org/hub>
TensorFlow Model Garden
<https://github.com/tensorflow/models/tree/master/official>

Trend: AI governance

- ¹¹ Explainable AI (XAI)
<https://www.infosys.com/services/incubating-emerging-technologies/documents/explainable-ai.pdf>

Advisory Council

Mohammed Rafee Tarafdar

SVP and Unit Technology Officer

Prasad Joshi

SVP, Emerging Technology Solutions

Balakrishna DR

SVP, Service Offering Head –
Energy, Utilities, Communications & Services

Satish HC

EVP, Head Global Services –
Data and Analytics

Hasit Trivedi

AVP – AI & Automation

Venkata Seshu G (Seshu)

AVP – Data and Analytics

Dinesh HR

EVP, Head Global Services –
Enterprise Package Application Services

Jasdeep Kaler

AVP, Global Product Head – NIA

John Gikopoulos

Partner – AI & Automation Practice

Shyam Kumar Doddavula

AVP, Emerging Technology Solutions

Thirumala A

VP – Education, Training and Assessment

Contributors

Aayush Agarwal

Allahbaksh Mohammedali Asadullah

Amit Gaonkar

Amit Saxena

Amitabh Manu

Chandra Sekhar Y

Divya Gopal Kalyani

Dr. Puranjoy Bhattacharya

Dr. Ravi Kumar G V V

Dr. Ravi Prakash G

Dr. Jithesh Sathyan

Kamalkumar Rathinasamy

Manjunath Kukkuru

Mohammed Rafee Tarafdar

Peeyush Singhal

Rajeev Nayar

Sanjay Mohan

Shashidhar Ramakrishnaiah

Shyam Kumar Doddavula

Sudhanshu Hate (primary author)

Swaminathan Natarajan

Thiruvengadam S

Venkata Lakshminarayana Indraganti

Venkata Seshu G (Seshu)

Vijayaraghavan Varadharajan

Vimal Balajee Viswanathan

Vittal Setty

Vivek Nagpal

Producer

Ramesh N

Infosys Knowledge Institute
ramesh_n03@infosys.com

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision-making on critical business and technology issues.

To view our research, visit Infosys Knowledge Institute at infosys.com/IKI.

For more information, contact askus@infosys.com



© 2020 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.

