

# AI-Powered NLP: The Evolution of Machine Intelligence from Machine Learning

BY **TUHIN CHATTOPADHYAY, PH.D.**  
BUSINESS ANALYTICS EVANGELIST

## QUICK VIEW

- 01** Classical machine learning techniques are used for text mining to accomplish sentiment analysis, topic modelling, TF-IDF, NER, etc.
- 02** With the advent of deep learning techniques, ML objectives like automated real-time question-answering, emotional connotation, fighting spam, machine translation, summarization, and information extraction are achieved.
- 03** Word embeddings, recurrent neural networks, and long short-term memory (LSTM) are used for content creation in author's style.

This article will illustrate the transition of the NLP landscape from a machine learning paradigm to the realm of machine intelligence and walk the readers through a few critical applications along with their underlying algorithms. Nav Gill's [blog](#) on the stages of AI and their role in NLP presents a good overview of the subject. A number of research papers have also been published to explain how to take traditional ML algorithms to the next level. Traditionally, classical machine learning techniques like support vector machines (SVM), neural networks, naïve Bayes, Bayesian networks, Latent Dirichlet Allocation (LDA), etc. are used for text mining to accomplish sentiment analysis, topic modelling, TF-IDF, NER, etc.

However, with the advent of open-source APIs like [TensorFlow](#), Stanford's [CoreNLP](#) suite, Berkeley AI Research's (BAIR) [Caffe](#), [Theano](#), [Torch](#), Microsoft's [Cognitive Toolkit](#) (CNTK), and licenced APIs like [api.ai](#), IBM's [Watson Conversation](#), [Amazon Lex](#), Microsoft's [Cognitive Services](#) APIs for speech (Translator Speech API, Speaker Recognition API, etc.), and language (Linguistic Analysis API, Translator Text API etc.), classical text mining algorithms have evolved into deep learning NLP architectures like recurrent and recursive neural networks. Google Cloud, through its [Natural Language API](#) (REST), offers sentiment analysis, entity analysis, entity sentiment analysis,

syntactic analysis, and content classification. Before diving further into the underlying deep learning algorithms, let's take a look at some of the interesting applications that AI contributes to the field of NLP.

To start with the craziest news, artificial intelligence is writing the sixth book of *A Song of Ice and Fire*. Software engineer [Zack Thoutt](#) is using a recurrent neural network to help wrap up George R. R. Martin's epic saga. [Emma](#), created by Professor Aleksandr Marchenko, is an AI bot for checking plagiarism that amalgamates NLP, machine learning, and stylometry. It helps in defining the authorship of write-up by studying the way people write. [Android Oreo](#) has the ability to recognize text as an address, email ID, phone number, URL, etc. and take the intended action intelligently. The smart text selection feature uses AI to recognize commonly copied words as a URL or business name. IBM Watson Developer Cloud's [Tone Analyzer](#) is capable of extracting the tone of any documents like tweets, online reviews, email messages, interviews, etc. The analysis output is a dashboard with visualizations of the presence of multiple emotions (anger, disgust, fear, joy, sadness), language style (analytical, confident, tentative), and social tendencies (openness, conscientiousness, extraversion, agreeableness, emotional range). The tool also provides sentence level analysis to identify the specific components of emotions, language style, and social tendencies embedded in each sentence.

[ZeroFox](#) is leveraging AI on NLP to bust Twitter's spam bot problem and protect social and digital platforms for

enterprises. [Google Brain](#) is conducting extensive research on understanding [natural language](#), and came up with unique solutions like autocomplete suggestions, autocomplete for doodles, and automatically answered e-mails, as well as the [RankBrain](#) algorithm to transform Google search. Google's [Neural Machine Translation](#) reduces translation errors by an average of 60% compared to Google's older phrase-based system. Quora conducted a Kaggle competition to detect duplicate questions where the modellers reach 90% accuracy. Last but not least, seamless question-answering is accomplished through a number of artificially intelligent natural language processors like Amazon's [Alexa Voice Service](#) (AVS), Lex, and [Polly](#), along with [api.ai](#), [archie.ai](#), etc. that can be embedded in devices like Echo and leveraged for virtual assistance through chatbots.

---

“While the focus of ML is natural language understanding (NLU), MI is geared up for natural language generation (NLG) that involves text planning, sentence planning, and text realization.”

---

Thus, the shift in gears from machine learning to machine intelligence is achieved through automated real-time question-answering, emotional analysis, spam prevention, machine translation, summarization, and information extraction. While the focus of ML is natural language understanding (NLU), MI is geared up for natural language generation (NLG) that involves text planning, sentence planning, and text realization. Conventionally, [Markov chains](#) are used for text generation through the prediction of the next word from the current word. A classic example of a Markov chain is available at [SubredditSimulator](#).

However, with the advent of deep learning models, a number of experiments were conducted through embedded words and recurrent neural networks to generate text that can keep the style of the author intact. The same research organization, Indigo Research, published a [blog](#) recently that demonstrates the application of long short-

term memory (LSTM) in generating the text through “memories” of a priori information. A number of research and development initiatives are currently going on the artificial natural language processing to match the human processing of language and eventually improve it.

The [Stanford Question Answering Dataset \(SQuAD\)](#) is one such initiative, with 100,000+ question-answer pairs on 5222300+ articles which were also shared in a [Kaggle competition](#). [Dynamic Co-attention Network \(DCN\)](#), which combines a co-attention encoder with a dynamic pointing decoder, gained prominence as the highest performer (Exact Match 78.7 and F1 85.6) in SQuAD and in [automatically answering questions](#) about documents. Other applications of deep learning algorithms that generate machine intelligence in the NLP space include bidirectional long short-term memory (biLSTM) models for [non-factoid answer selection](#), convolutional neural networks (CNNs) for [sentence classification](#), recurrent neural networks for [word alignment models](#), word embeddings for speech recognition, and [recursive deep models](#) for [semantic compositionality](#). Yoav Goldberg's magnum opus and all the dedicated courses [Stanford, Oxford, and Cambridge] on the application of deep learning on NLP further bear testimony to the paradigm shift from ML to MI in the NLP space.

With the evolution of human civilization, technological advancements continue to complement the increasing demands of human life. Thus, the progression from machine learning to machine intelligence is completely in harmony with the direction and pace of the development of the human race. A few months ago, Nav Gill's [blog](#) on the stages of AI and their role in NLP observed that we have reached the stage of machine intelligence, and the next stage is machine consciousness. Of late, AI has created a lot of hype by some who see it as the greatest risk to civilization. However, like any technology, AI can do more good for society than harm — when used correctly. Instead of the predicted cause of the apocalypse, AI may turn out to be the salvation of civilization with a bouquet of benefits, from [early cancer detection](#) to [better farming](#).

**Tuhin Chattopadhyay** is a business analytics and data science thought leader. He was awarded Analytics and Insight Leader of the Year in 2017 by KamiKaze B2B Media and was featured in India's Top 10 Data Scientists 2016 by Analytics India Magazine. Tuhin spent the first ten years of his career in teaching business statistics, research, and analytics at a number of reputed schools. Currently, Tuhin works as Associate Director at The Nielsen Company and is responsible for providing a full suite of analytics consultancy services to meet the evolving needs of the industry. Interested readers may browse his [website](#) for a full profile.

