

# OVERVIEW OF NATURAL LANGUAGE PROCESSING (NLP) TECHNOLOGIES

---

Presented by: WTI  
[www.wti-solutions.com](http://www.wti-solutions.com)  
703.286.2416



# LEGAL DISCLAIMER

The entire contents of this informational publication is protected by the copyright laws of the United States and other jurisdictions. You may print a copy of any part of this publication for your own personal, noncommercial use, but you may not copy any part of the publication for any other purposes, nor may you modify any part of the publication. Inclusion of any part of the content of this publication in another work, whether in printed or electronic, or other form, or inclusion of any part hereof in another publication or web site by linking, framing, or otherwise without the express written permission of WTI is hereby prohibited.

WTI's informational publications are made available for educational purposes only as well as to give you general information and a general understanding of technology, not to provide technical advice. By reading our informational publications you understand that there is no contractual relationship created between you and WTI. Although the information in our informational publications is intended to be current and accurate, the information presented herein may not reflect the most current technical developments. These materials may be changed, improved, or updated without notice. WTI is not responsible for any errors or omissions in the content of this publication or for damages arising from the use or performance of this publication under any circumstances. We encourage you to contact us or other technology companies for specific technical advice as to your particular matter.

# WHAT IS NLP AND WHY IS IT IMPORTANT?

- **Definition** - Natural language processing (NLP) is the merging of computer processing and computational linguistics with artificial intelligence to create functions that allow the processing of large amounts of human produced data (written or spoken) into machine understandable data sets for analytics.
- **Importance** - The challenges it attempts to solve are natural language understanding, natural language generation, connecting language to machine perception and creating dialog systems.

# NLP ORIGINS

- First established as a field in “Computing Machinery and Intelligence” by Alan Turing in 1950, who outlined the possibilities of what would be known as the Turing Test - in which a computer could understand questions and respond well enough to fool a human tester in believing they were speaking to a live human rather than a machine.
- Work began to create systems to automatically translate documents between different languages.
- A decade later the concept of the technology would be popularized by television programs such as “Star Trek” and movies like 2001: A Space Odyssey, which showed computers understanding and replying in natural fashion.<sup>1</sup>

<sup>1</sup> Deangelis, Stephen F. (Feb 2014) The Growing Importance of Natural Language Processing. Retrieved from <https://www.wired.com/insights/2014/02/growing-importance-natural-language-processing/>

# NLP EVOLUTION

While a popular science fiction concept, the technology failed to gain real advancement for many decades. Concerned by slow progress in 1964, U.S. Government sponsors commissioned the Automatic Language Processing Advisory Committee (ALPAC) which concluded in 1966 and reported, shortsightedly, that it was too slow and too costly compared to human workers so “there is no immediate or predictable prospect of useful machine Translation”.<sup>2</sup> This report chilled research in human language machine learning for decades.

<sup>2</sup> Hutchins, John (Nov 2005) The History of Machine Learning in a Nutshell. Retrieved from <http://www.hutchinsweb.me.uk/Nutshell-2005.pdf>

# NLP REBIRTH

The late 1990s saw the rebirth of NLP with the rise of the World Wide Web and the need for real-time information processing became a mass market need. This pushed interest in NLP from a mostly research topic to a focus on practical applications<sup>3</sup> such as automated translation and search engines.

Around the same time on the research side, several breakthroughs combining old rule-based systems with newer statistical and example-based systems increased the complexity that NLP technologies may be able to handle.

<sup>3</sup> Hutchins, John (Nov 2005) The History of Machine Learning in a Nutshell. Retrieved from <http://www.hutchinsweb.me.uk/Nutshell-2005.pdf>

# NLP IN THE NEWS

## IBM Watson Wins Jeopardy!

Watson, a primarily NLP engine, defeats the top human champions in 2011. Starting as a stand-alone supercomputer, it has since evolved into a powerful cloud-based engine capable of outperforming doctors at medical diagnosis or legal paperwork processing.<sup>4</sup>

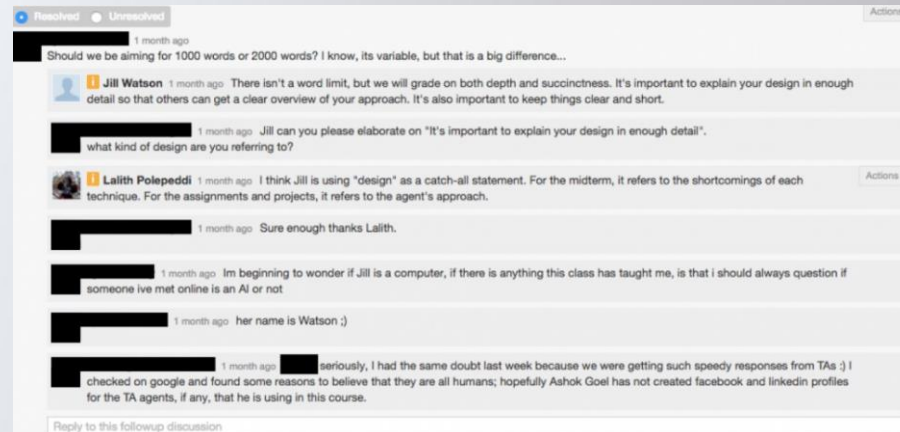


<sup>4</sup> Captain, Sean (Jan 2017) Can IBM's Watson do it all? Retrieved from <https://www.fastcompany.com/3065339/can-ibms-watson-do-it-all>

# NLP IN THE NEWS

## NLP Teaching Assistant Fools Students

Utilizing the later cloud-based Watson technology, an artificial intelligence professor at Georgia Tech created a virtual teaching assistant to help respond to the over 10,000 inquiries received from students. It did a good enough job that most did not realize their answers came from a computer.<sup>5</sup>



<sup>5</sup> Bussing, Kim (Sept 2016) Georgia Tech's Teaching Assistant "Jill Watson" Turns Out To Be A Robot! Retrieved from <https://www.dogonews.com/2016/9/10/georgia-techs-teaching-assistant-jill-watson-turns-out-to-be-a-robot>



# NLP IN THE NEWS

## Advertisements Activate Home Devices

Taking advantage of the rise of home assistants from Google and Amazon, advertisers created commercials intended to activate Google home devices and use them to continue giving product information beyond the commercial's length. Google quickly updated the devices to prevent the specific ad's command. Prevention of unwanted activations remains an ongoing challenge.<sup>6</sup>

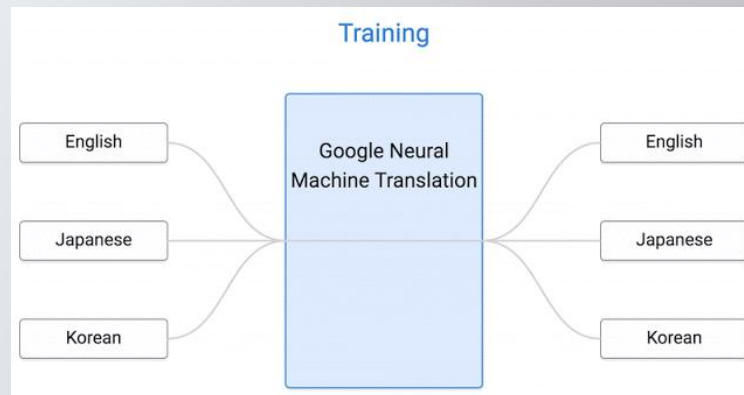


<sup>6</sup> Maheshwari, Sapna (Apr 2017) Burger King 'O.K. Google' Ad Doesn't Seem O.K. With Google. Retrieved from <https://www.nytimes.com/2017/04/12/business/burger-king-tv-ad-google-home.html>

# NLP IN THE NEWS

## Google Translate Invents its Own Language

An exaggerated headline to explain a still-impressive achievement in NLP, the system behind Google Translate was testing a new “neural machine translation” and the system, unprompted, determined the best way translate two languages that did not have direct correlations was to create an “interlingua” using data from a third language to assist in figuring out the correct translation. <sup>7</sup>



<sup>7</sup> Reback, Gedalyah (Jan 2017) No, Google Translate did not invent its own language called ‘interlingua’. Retrieved from <https://www.geektime.com/2017/01/23/no-google-translate-did-not-invent-its-own-language-called-interlingua/>

# NLP EVERYDAY

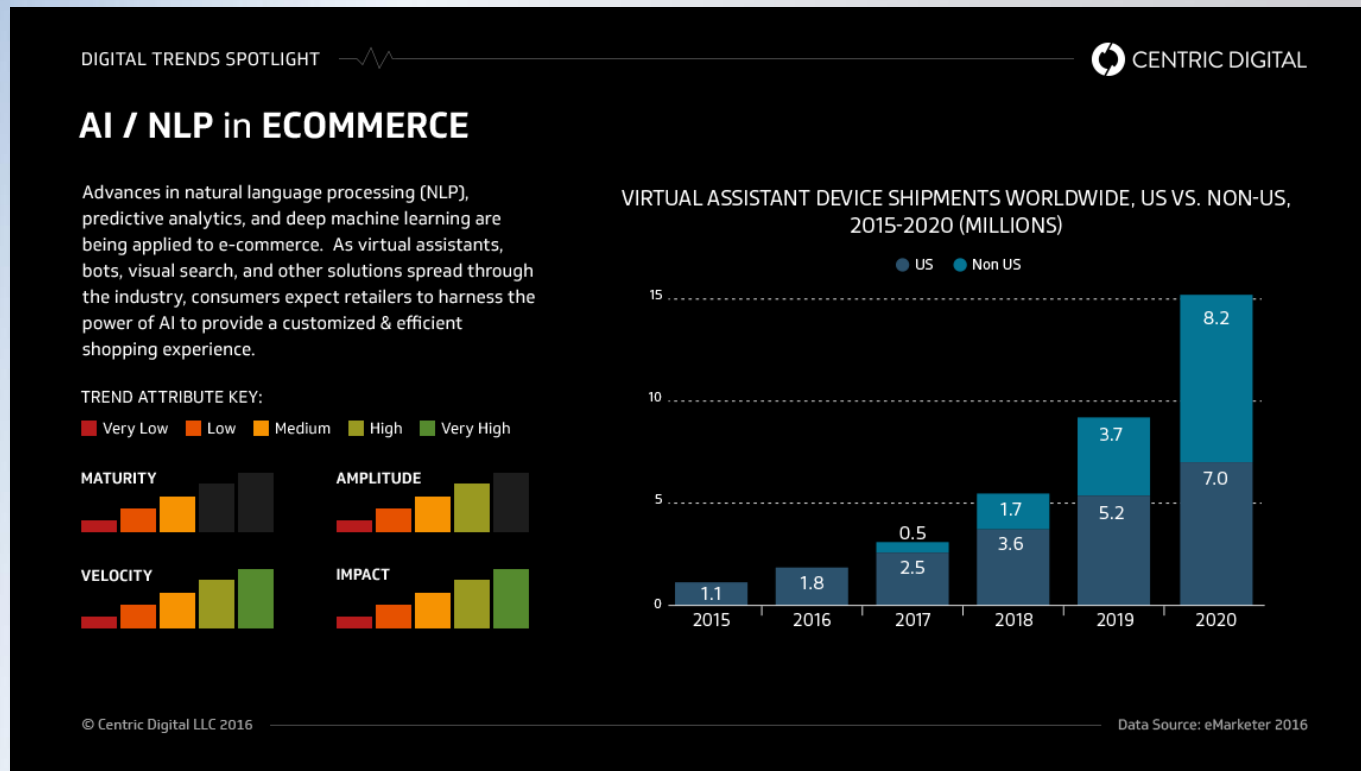
NLP technologies are also moving beyond large dataset crunching and becoming part of our everyday lives. For example, NLP is used in everyday interfaces such as:

- Apple's Siri
- Google's "Ok Google"
- Automated support chat boxes
- Automated phone answering services
- Automated journalism articles
- Optical Character Recognition technologies<sup>8</sup>

<sup>8</sup> "Outlook on Artificial Intelligence in the Enterprise 2016", Narrative Science in partnership with National Business Research Institute, Page 6, retrieved from <http://www.datascienceassn.org>

# NLP FOR VIRTUAL ASSISTANTS

As the usefulness of NLP functions increase, consumer demand for virtual assistants is expected to rise sharply.



<sup>9</sup> <https://centricdigital.com/blog/artificial-intelligence/how-natural-language-process-artificial-intelligence-are-changing-e-commerce/>

# HOW DOES NLP DIFFER?

Use of NLP is different from prior processes for older systems in which the exact rules had to be entered to create the program. NLP systems are trained using machine learning to analyze example data sets and make inferences. Examples include:

- Summarizing blocks of text to extract the most important and central ideas while ignoring irrelevant information.
- Creating a chat bot that uses Point-of-Speech tagging.
- Automatically generating keyword tags from content that leverages latent Dirichlet allocation (LDA), a technique that discovers topics contained within a body of text.
- Using Sentiment Analysis to identify the sentiment of a string of text, from very negative, to neutral to very positive.<sup>10</sup>

<sup>10</sup> Kiser, Matt (Aug 2016) Introduction to Natural Language Processing (NLP) 2016. Retrieved from <https://blog.algorithmia.com/introduction-natural-language-processing-nlp/>.

# WHY DOES NLP MATTER TO GOVERNMENT?

Government organizations have the need for rapid processing of large text datasets. NLP can reduce the workload and speed response. Examples include:

- Civic Government - Sorting public comments to understand policy response.
- Emergency Services - Insurance and disaster response.
- Healthcare - Paperwork processing and diagnostic assistance.
- Law Enforcement – Crime solving and prevention by analyzing multiple case files for patterns and connections.<sup>11</sup>

<sup>11</sup> Murrow, Brian (Mar 2014) Extracting Meaningful Analysis from Data Using Natural Language Processing. Retrieved from <https://www.ibm.com/blogs/insights-on-business/government/extracting-meaningful-analysis-from-data-using-natural-language-processing/>

# EXAMPLES OF WTI NLP WORK

WTI works with Government agencies to provide innovative NLP solutions that:

- Extract data from the web and other sources in order to persist the data.
- Integrate commercial NLP products and custom code to extract “entities” from text.
- Classify entities using defined ontologies that describe key Government topic domains.
- Execute processes to create relationships between extracted entities.
- Leverage RDF and other object stores to store data in accordance with defined ontologies.
- Enable queries via various COTS, GOTS, and custom tools against the results of NLP.

# CONTACT INFORMATION

WTI is a CMMI®-DEV Level 3 appraised and SBA 8(a) Certified, Economically Disadvantaged Woman-Owned Small Business based in the National Capital Region, committed to helping the Government and industry harness the power of actionable information through intelligence and IT solutions.

WTI's work in NLP spans multiple Federal Government executive departments and a number of independent Federal agencies. To find out more, contact: [info@wti-solutions.com](mailto:info@wti-solutions.com).