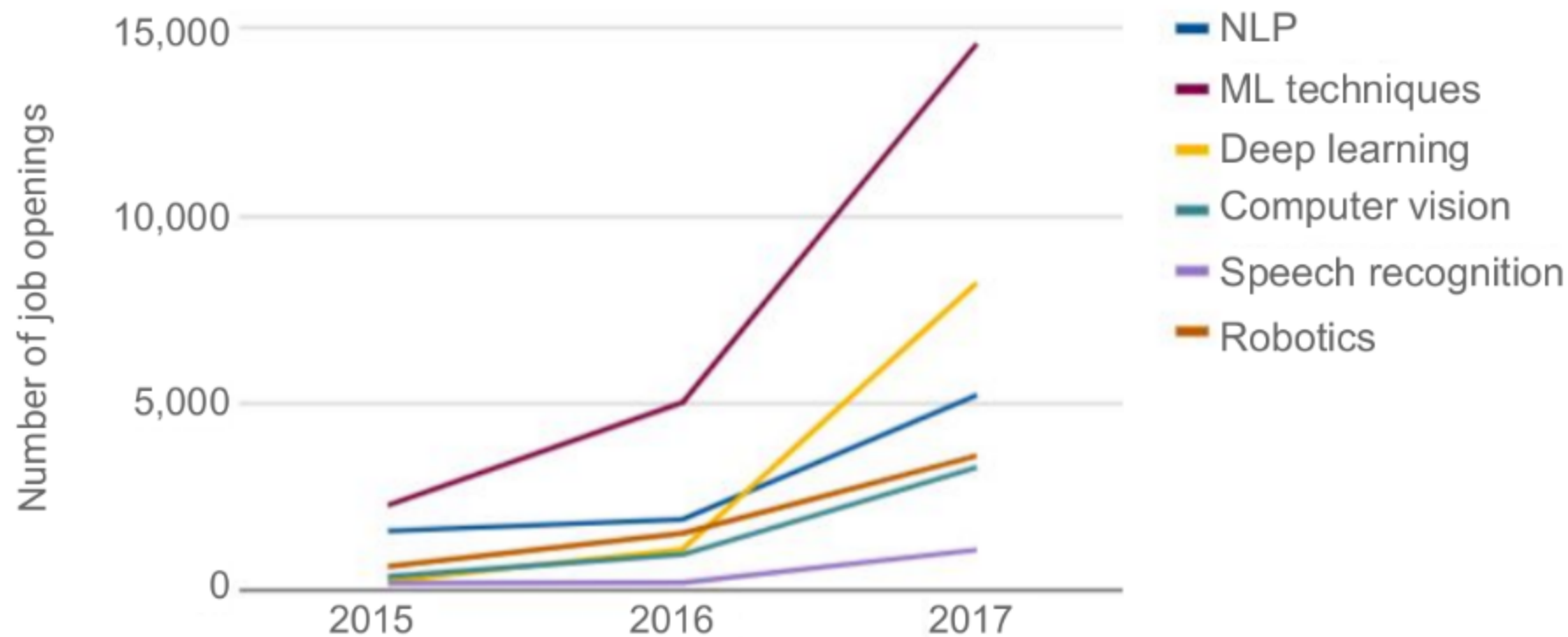


# Artificial Intelligence

NLP and Deep Learning

## Job openings by AI skills required (2015 – 2017)

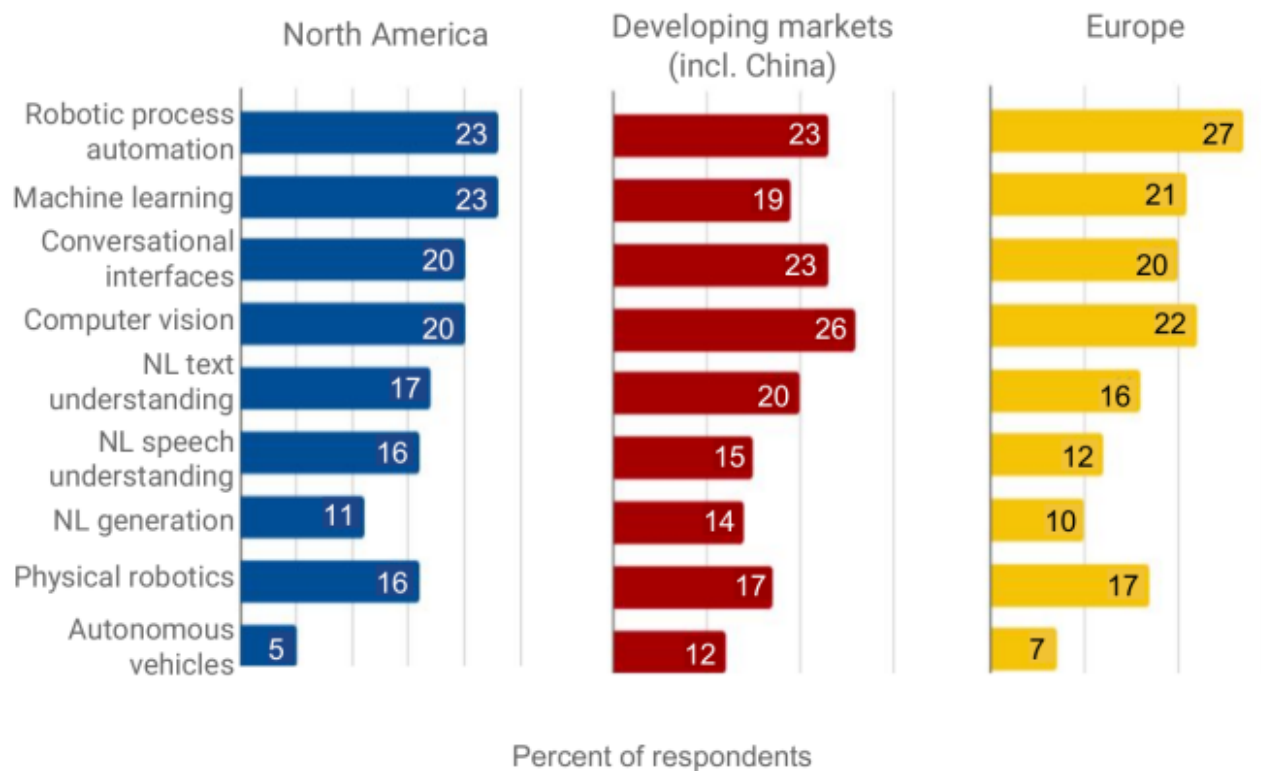
Source: Monster.com



Source: [2018 AI Index report](#)

## Capabilities embedded in at least one company function (2018)

Source: McKinsey & Company



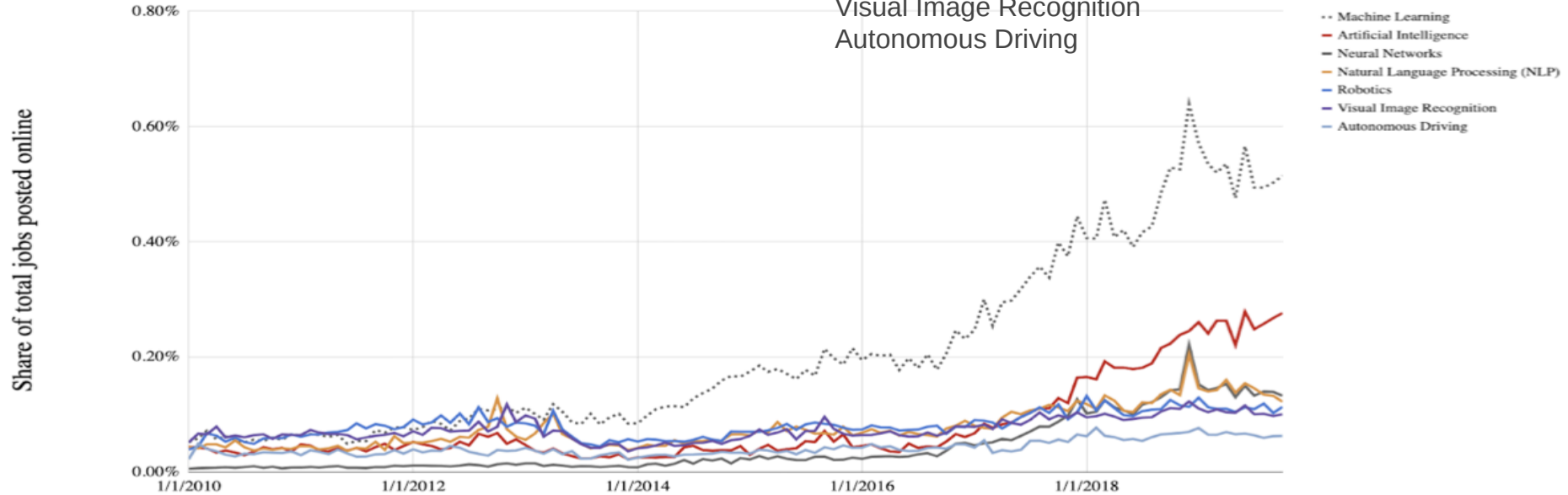
Note: The size of each bar is relative to the capabilities within each region; North America: N = 479; Developing markets (incl. China): N = 189 (China N = 35); Europe: N = 803

Source: [2018 AI Index report](#)

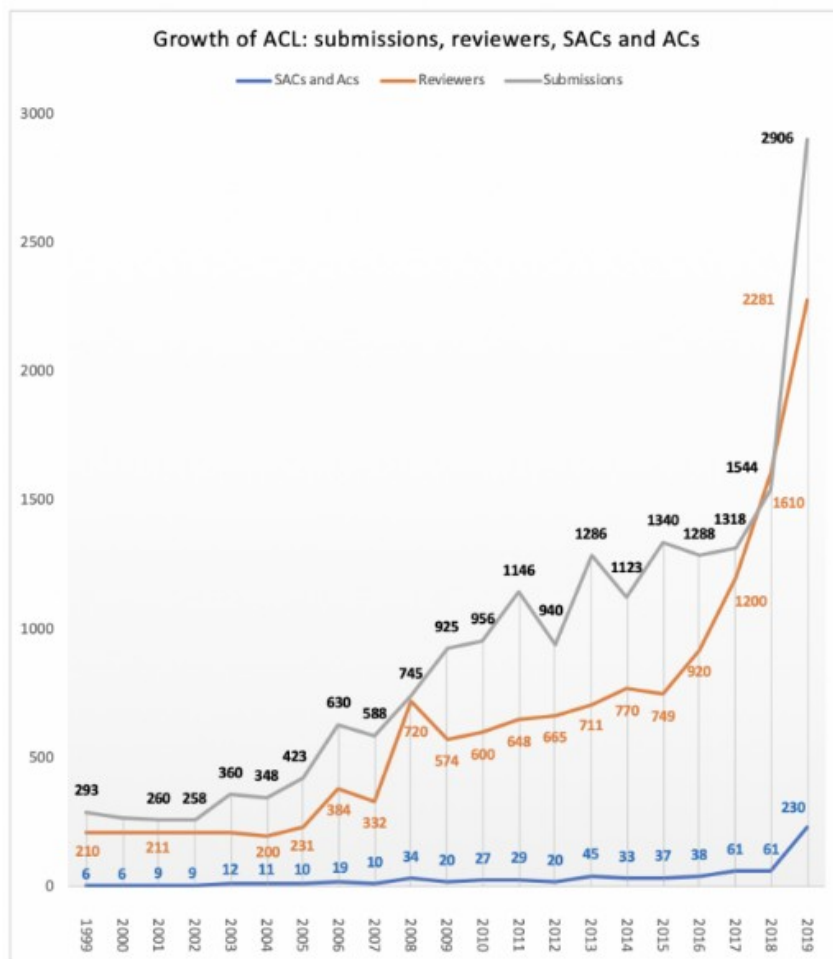
Machine Learning  
Artificial Intelligence  
Neural Networks  
**Natural Language Processing**  
Robotics  
Visual Image Recognition  
Autonomous Driving

Share of Total Online Job Postings, USA, 2010-2019 monthly

Source: BurningGlass, 2019.



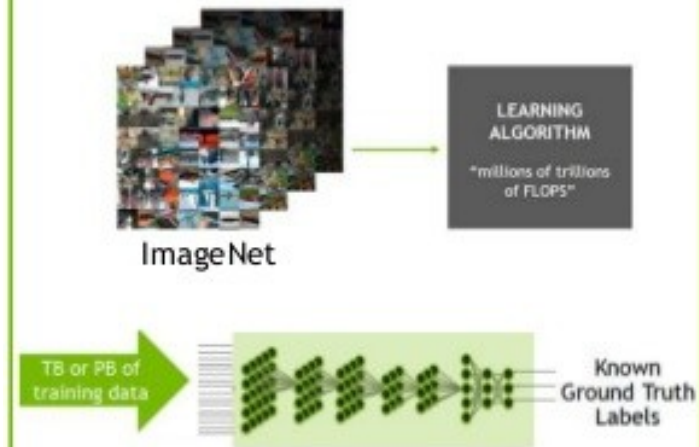
Source: [2019 AI Index report](#)



Source: <http://acl2019pcbolog.fileli.unipi.it/>

# DEEP LEARNING - A NEW COMPUTING MODEL

## “Training”

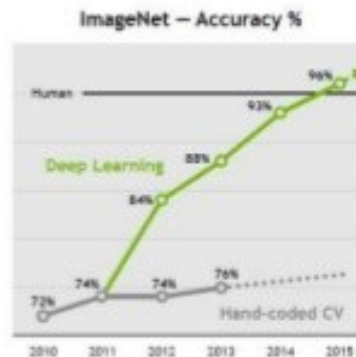


## “Inference”



“cat”

## “SUPERHUMAN” RESULTS SPARK HYPERSCALE ADOPTION



From Andy Steinbach (NVIDIA)

# Superhuman results

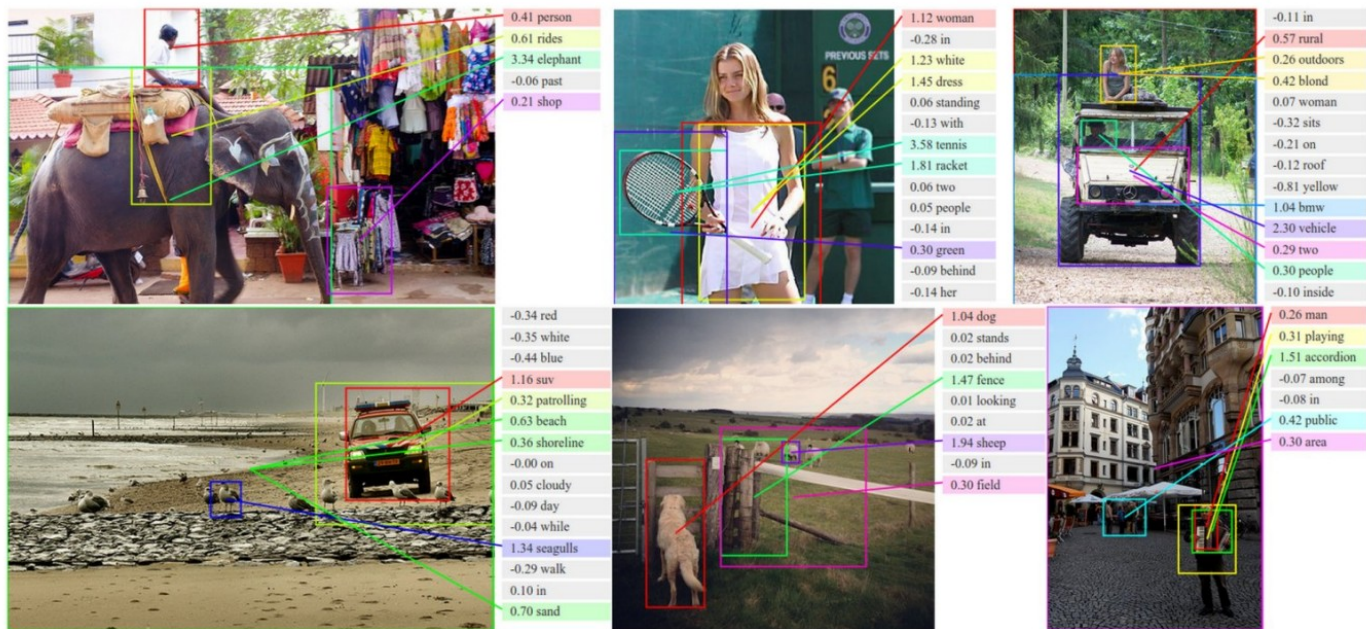


Figure 5. Example alignments predicted by our model. For every test image above, we retrieve the most compatible test sentence and visualize the highest-scoring region for each word (before MRF smoothing described in Section 3.1.4) and the associated scores ( $v_i^T s_t$ ). We hide the alignments of low-scoring words to reduce clutter. We assign each region an arbitrary color.

## Deep visual-semantic alignments for generating image descriptions (2014)

A Karpathy, L Fei-Fei



# Superhuman results

## StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks (2016)

Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, Dimitris Metaxas

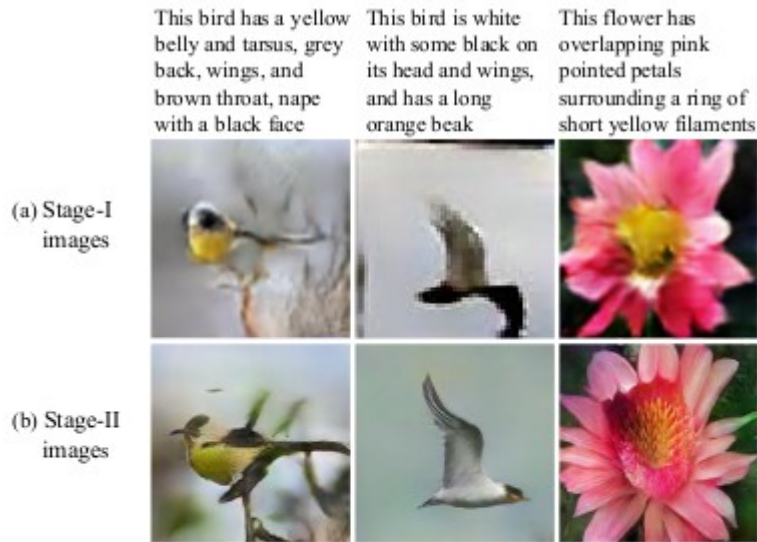


Figure 1. Photo-realistic images generated by our StackGAN from unseen text descriptions. Descriptions for birds and flowers are from CUB [32] and Oxford-102 [18] datasets, respectively. (a) Given text descriptions, Stage-I of StackGAN sketches rough shapes and basic colors of objects, yielding low resolution images. (b) Stage-II of StackGAN takes Stage-I results and text descriptions as inputs, and generates high resolution images with photo-realistic details.



# Superhuman results



## **LipNet: Sentence Level Lipreading (2016)**

Yannis M. Assael, Brendan Shillingford, Shimon Whiteson, Nando de Freitas

# Superhuman results

## Zero-Shot Text-to-Image Generation (2021)

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever

<https://openai.com/blog/dall-e/>

TEXT PROMPT

an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES



Edit prompt or view more images ↴

TEXT PROMPT

an armchair in the shape of an avocado . . .

AI-GENERATED IMAGES



Edit prompt or view more images ↴

TEXT PROMPT

a store front that has the word 'openai' written on it . . .

AI-GENERATED IMAGES

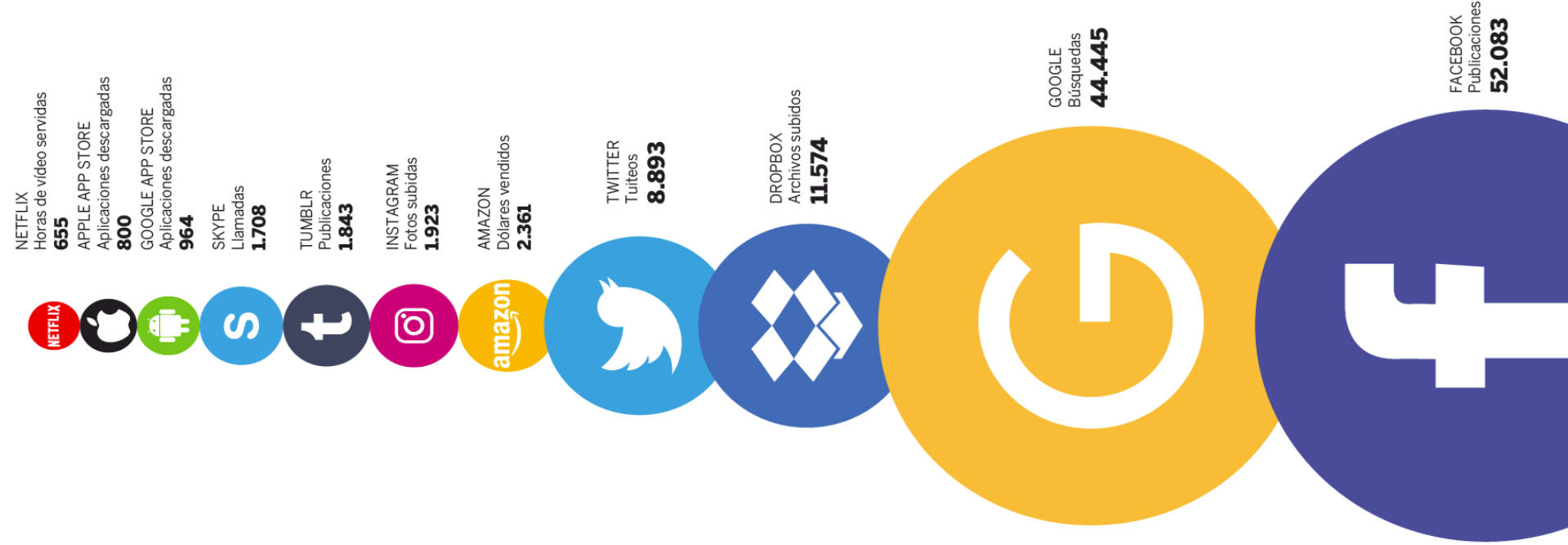


Edit prompt or view more images ↴

# Natural Language Processing

- **Unstructured** digital content accounts for **80%** of all information [ [White paper IDC 2014](#)] ...
- Usually in the form of **texts** (also audio, video, etc.) and documents in multiple **languages** ...
- Only appropriate **NLP tools** can access this wealth of knowledge ...
- NLP among the **top** 10 strategic technology trends for 2019 according to [Gartner](#)

# What happens in Internet every **second**? (july 2015)



# What happens in Internet every **second**? (july 2015)



## ... not only from Social Media

- LexisNexis receives **daily** 1.5M news.
- ?M judicial sentences, transcriptions ...
- ?M Electronic Health Records (EHR) ...
- ?M Patents ...
- ... all kinds of e-documents ...
- ... and only **appropriate techniques** can handle all this digital data
- After finishing the degree, most computer science students **do not know** how to process most of the digital information.
  - Final grade works, NLP and DL courses, LAP Master degree, PhD ...
  - New AI degree ...

# HiTZ Center

- Promote **training**, **research** and **technological transfer** in language technologies
- **65** members
- **2** research groups

ahō LAB

ixa





# HiTZ center

National and international **reference** in LT

- > **1.5** million euros annual budget
- > **500** research papers
- > **100 research** and **transfer** projects
- > **30** companies and institutions use our products
- > **16** awards (3 Google, 1 BBVA, 5 Best paper, Kaggle COVID-19, CLEF ICD, **Computer Science National prize**, etc.)



# HiTZ Center

**L&CT Erasmus Mundus** Máster

ENRICH **European Training Network**

**Doctoral** program

> **60** Doctors

> **25** doctoral student

continuous training:

[Deep Learning for NLP](#)

[Language-centered AI Applications](#)



Erasmus  
Mundus

*Erasmus Mundus*

*LCT -- European Masters Program in  
Language & Communication Technology*



ixa

Contents Instructors Practical details Previous editions

# Deep Learning for Natural Language Processing (5th edition)

## 35-hour, 3-week winter course

From January 27th to February 13th 2020 in San Sebastian

[Register](#)

[http://ixa2.si.ehu.es/deep\\_learning\\_seminar/](http://ixa2.si.ehu.es/deep_learning_seminar/)

Deep Learning neural network models have been successfully applied to natural language processing, and are now changing radically how we interact with machines (Siri, Amazon Alexa, Google Home, Skype translator, Google Translate, or the Google search engine). These models are able to infer a continuous representation for words and sentences, instead of using hand-engineered features as in other machine learning approaches. The seminar will introduce the main deep learning models used in natural language processing, allowing the attendees to gain hands-on understanding and implementation of them in Tensorflow.



**HiTZ**

Hizkuntza Teknologiako Zentroa  
Basque Center for Language Technology

[Contents](#) [Instructors](#) [Practical details](#)

# Introduction to Language Technology Applications (2nd edition)

## 25 hour, 5 week winter course

February 2nd to March 9th 2022, Donostia-San Sebastian

*NOTE: Due to the pandemic, online and onsite assistance available*

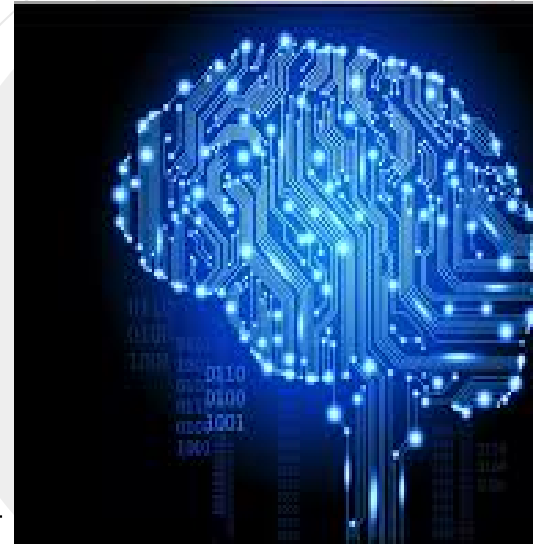
Registration Open!

See also our sister course "[Deep Learning for Natural Language Processing](#)"

<https://www.ixaeus/iltapp/>

# New AI paradigm

- **Deep Learning**, the neural revolution
  - Solution for **Big Data** tasks
  - Requires **HPC** (supercomputing)
- **Superior** to humans in many tasks
- 8 editions of the Deep Learning **training course** (more than 300 people trained)
- **New** Language-centered AI [training course](#)
- The HiTZ center is **leading** the development of Deep Learning applications
- Translation without bilingual data (**Science** magazine)



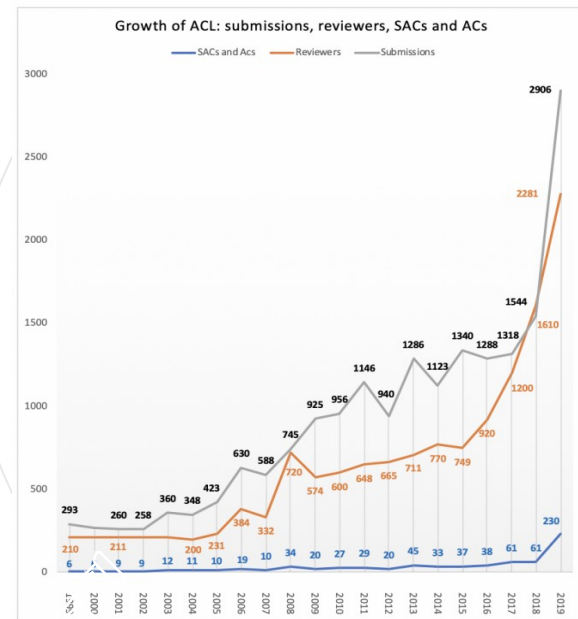


# Language-centered AI

**80%** of the digital information is **unstructured Text** (multilingual), audio, video

*Natural Language Processing* is the area of **Artificial Intelligence** that deals with written and spoken language

- Machine Translation
- Information extraction and retrieval
- Speech Recognition and Synthesis
- Dialogue systems (chatbots)



# Research policy

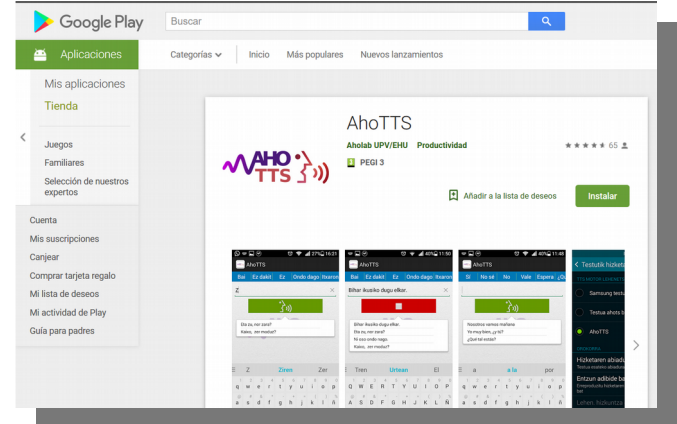
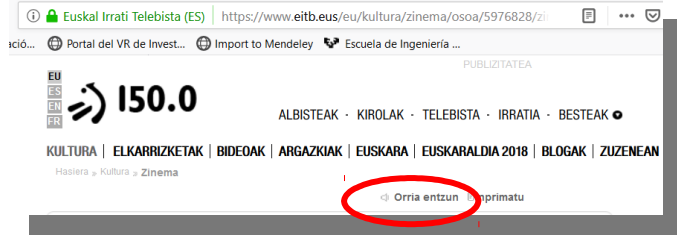
- **Advisors** of Spanish LT plan since its inception
- **Technology transfer** and **innovation**
  - **Health** and **justice** vertical, among other
  - Applications for **public administration**
- Authors of the Basque **TL report**
- [National AI Plan](#) & [European AI Plan](#)
- **European Language Equality** project: Roadmap and Agenda to achieve digital language equality in Europe by 2030.





# Automatic speech and audio processing Voice generation.

Developments for Euskara Improved communication skills of people with oral disabilities (ALS, laryngectomized) Speech recognition and information extraction from voice and audio (Machine Listening)



## ¿Qué es AhoMyTTS?

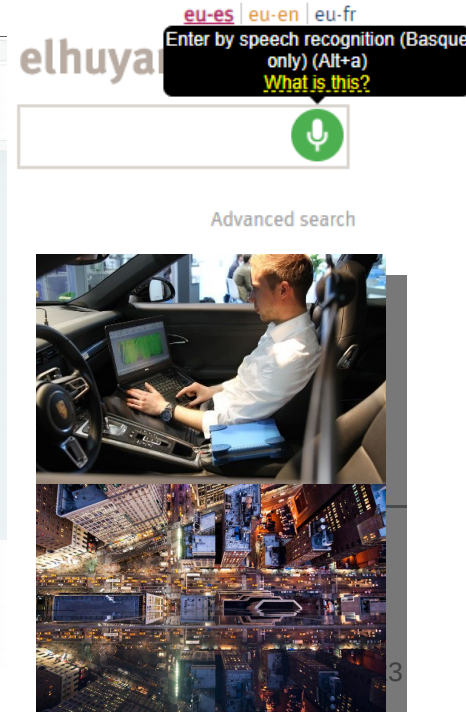
AhoMyTTS es una iniciativa del **Grupo Aholab** de la UPV/EHU con el fin de poder proporcionar **Sistemas de Conversión de Texto en Voz** personalizados a personas con

## Banco de Voces Sintéticas

Desde aquí podrás escuchar una selección de las disponibles en el "Banco de Voces Sintéticas" y "Tu voz" si ya la has donado.

## Descargo de responsabilidad

Para participar en este proyecto, el usuario **consiente expresamente** aceptar las presentes condiciones.



## Machine Translation

Pioneers in the use of Deep Learning for Basque, improvements in specific domains

Discovery: it is possible to translate between languages without bilingual resources

Science Home News Journals Topics Careers

f 317  
t 15  
in 193



Computers might soon translate between many more languages.

iStock.com/Lightcome

### Artificial intelligence goes bilingual—without a dictionary

By Matthew Hutson | Nov. 28, 2017, 4:30 PM



encial

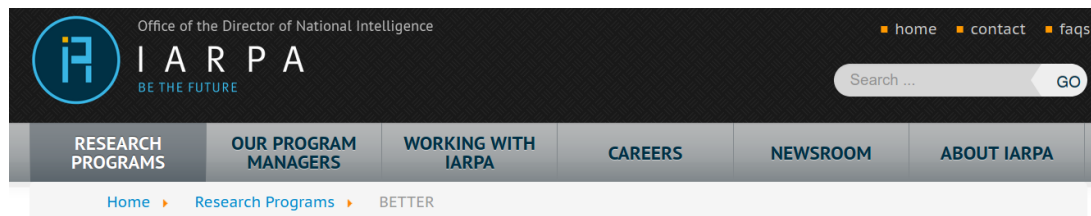
os vascos han creado un traductor para hacer sombra a ...

ingenieros vascos han creado un traductor para hacer sombra  
s de Tecnología. Un equipo de investigadores de la Universi...

## Content Management

Every day, 3 trillion words are published. Where can I find the information I need?

Information Extraction (IE) and Information Retrieval (IR)



Office of the Director of National Intelligence

**IARPA**  
BE THE FUTURE

home contact faqs

Search ... GO

RESEARCH PROGRAMS | OUR PROGRAM MANAGERS | WORKING WITH IARPA | CAREERS | NEWSROOM | ABOUT IARPA

Home > Research Programs > BETTER

## Better Extraction from Text Towards Enhanced Retrieval (BETTER)



The Better Extraction from Text Towards Enhanced Retrieval (BETTER) Program will

develop methods for extracting increasingly fine-grained semantic information, with a focus of events in the form of who-did-what-to-whom-when-where, across multiple languages and problem domains. This extracted information will be applied to an information retrieval task. An additional area of focus is human-in-the-loop computation. Performer systems will need the ability to incorporate human judgments for metrics such as relevancy and accuracy of extracted or retrieved information.

### Program Manager

[Carl Rubino](#)

### Program Information

[IARPA-BAA-18-05](#)

### Research Area(s)

- Human language technology
- Natural language processing
- Information extraction
- Information retrieval
- Active learning
- Multilingual processing



**deia**

INICIO | BIZKAIA | ATHLETIC | POLÍTICA | SOCIEDAD | DEPORTES | CULTURA Y OCM

Euskadi | Estado - Mundo | Obituarios | Historias de los vascos | Euskadi y Bizkaia en datos

Inicio > Sociedad > Euskadi

50.000 DOLARREKO DIRU LAGUNTZA JASOKO DU

### Eneko Agirre EHUKo irakasleak Google Research Awards sarietako bat jaso du

Guztira, Googlek 151 proposamen saritu ditu, 55 herrialdetako 350 unibertsitatek bidalitakoak

EP - 2016(e)ko martxoaren 18a, ostirala - 12:22etan eguneratu

comental | G+ | Tweet | Atsegin dut | Elkarbanatu



# Media and Social Networks

Positive and negative tweets (DSS2016), real-time monitoring and analysis.  
Research on fine tuning positive and negative aspects. Fake news. Multilingual.

DONOSTIA / SAN SEBASTIÁN 2016 EUROPAKO KULTUR HIRIBURUA  EU | ES | EN | FR | [buletina](#)

## Behagunea

Orokorra  Bakea  Ahotsak  Bizitza  Denak

[Etiketa lainoa](#) [Gaiak](#) [Egileak](#)



**Codefest** **Hirikilabs** **Kalebegiak** **Donostiapedia**  
**Zubideak** **San Sebastian 2016** **Euro-Diálogos**  
**2016 Bidea** **Euro-Dialogues**  
**Irekiera** **Biziz** **Donostia 2016** **Olatu** **Talka** **Tosta**  
**Hitzargiak** **Itsasfest** **Gaur** **Konstelazioak** **Tira** **Birak**  
**WhatsART** **Stop\_War** **Festibala** **Emusik** **La Publika** **Euroku**  
**Feministaldia** **Lecturas 2016** **Europa** **Transit** **Hibrilaldiak**  
**Rekording** **Ondarebideak** **Tratado De Paz** **Bake Ituna**



Sentiment	Count	Percentage
Positiiboak	12665	28%
Negatiiboak	4799	11%
Neutroak	27927	62%

[Aipamenak](#) ([Filtroak]: date: Duela hilabete bat)

Denak	Aipamen positiboak	Aipamen negatiiboak
<p>👤 <a href="#">MikelBolivar</a>: At the end... This is the end, my friend... of 2016... DSS2016 ?... Welcome 2017... best wishes... <a href="https://t.co/TcP3WPNeY5">https://t.co/TcP3WPNeY5</a> + 2016-12-31 23:33</p> <p>👤 <a href="#">GonzaloMartinez</a>: Re #Lecturas2016 <a href="https://t.co/DGcNZedyVM">https://t.co/DGcNZedyVM</a> + 2016-12-31 23:23</p> <p>👤 <a href="#">SincerelyWendy</a>: #WavesOfEnergy This chilling 'Star Wars blaster fire' sound coming from the Northern Lights has got everyone spooked <a href="https://t.co/xghfMsdR1">https://t.co/xghfMsdR1</a> + 2016-12-31 21:03</p> <p>👤 <a href="#">VeronicaBLH</a>: #Lecturas2016 23 novelas, 23 historias Dime quién soy Los besos en el pan El bar de las grandes esperanzas La trilogía de @DOLORESREDONDOM + 2016-12-31 20:19</p> <p>👤 <a href="#">Papurudia</a>: @DSS2016 #energiaolatuak @Papurudia @BAGERAEkartera @Tikiik Urte berri on eta 2017an <a href="https://t.co/TCp3WPNeY5">https://t.co/TCp3WPNeY5</a> + 2016-12-31 23:33</p> <p>👤 <a href="#">Papurudia</a>: @DSS2016 #energiaolatuak @Papurudia @BAGERAEkartera @Tikiik Urte berri on eta 2017an gehiago! <a href="https://t.co/NpeV40KxFP">https://t.co/NpeV40KxFP</a> + 2016-12-31 19:06</p> <p>📰 <a href="#">Diario Vasco</a>: Paya también ha querido destacar que "San Sebastián 2016 ha hecho tantas cosas que es imposible reducirlo en 63 minutos -duración del documental-. Se ha puesto hincapié en el programa, es lo que más se puede ver, pero hay infinidad de labores de in... + 2016-12-31 17:59</p> <p>👤 <a href="#">ideaspalabras</a>: Mis favoritos de 2016 ? #lecturas2016 #lecturasrecomendadas #misfavoritos <a href="https://t.co/TCp3WPNeY5">https://t.co/TCp3WPNeY5</a> + 2016-12-31 17:59</p> <p>👤 <a href="#">xpsicic</a>: Lo mejor y lo peor que he leído en 2016. <a href="https://t.co/1lmKCXwSON">https://t.co/1lmKCXwSON</a> #Lecturas2016 #librosrecomendados + 2016-12-31 17:59</p> <p>👤 <a href="#">JazzBruno</a>: Año de lecturas aunque hicieran falta varias que no tienen registro <a href="https://t.co/1lmKCXwSON">https://t.co/1lmKCXwSON</a> via @goodreads #libros2016 #lecturas2016 + 2016-12-31 17:59</p> <p>👤 <a href="#">Angel_Barko</a>: Se acabó Donostia 2016 y la sensación de la ciudadanía es que ha pasado gloria, un fiasco, mi opinión. #Donostia2016 + 2016-12-31 17:59</p> <p>👤 <a href="#">ereitenk</a>: @bsarasola @gantzarain @hedoi_etxarte almandoz ormazabal rezola fernandez azkenean @DSS2016 kulturaz ta arteaz hausnartzeko balio duen + 2016-12-31 17:59</p>		

# Multilingual Chatbots and dialog

## FAQ Question Answering systems

Cuál es la política de devoluciones? 10:02 ✓

Puedes devolver tu pedido o parte del mismo. Ten en cuenta que dispones de 14 días hábiles desde el momento de la compra. 10:02

Quiero actualizar mis datos personales 10:02 ✓

Solo tienes que entrar aquí, editar los datos que quieras modificar y pinchar en Actualizar Datos. 10:03

Es seguro pagar online? 10:03 ✓

La seguridad y protección de tus datos es lo más importante para nosotros. 10:03

Qué métodos de pago aceptáis? 10:04 ✓

Puedes hacer el pago de tus compras mediante tarjeta de crédito o débito (Visa, Electron, MasterCard, Maestro, Euro6000, Telebanco) o a través de Paypal. 10:04

Message



How can I contact you? 15:21 ✓

You can contact us through email here, bark at us via chat, or text or call us at [855-944-2275](tel:855-944-2275) 15:21

Do you have trial offers? 15:22 ✓

Please note: we do not have "trial" offers. We do have a 1 month subscription plan option at sign-up if you want to sign up with a month-to-month commitment instead! 15:22

Do you offer any pay-as-you-go plans? 15:22 ✓

We do have a 1 month subscription plan option at sign-up if you want to sign up with a month-to-month commitment 15:23

Do you offer boxes for cats? 15:23 ✓

That is not currently an option that we offer. Sorry, kitties! We still love you. 15:24

Message





# Language-centered eHealth

**Basque** SNOMED CT and CIE-10-MC machine translation

**Spanish** Detection of adverse drug reactions and ICD-10 coding

**English** CORD-19 (COVID-19 Open Research Dataset Challenge), organized by Allen Institute for AI, Chan Zuckerberg Initiative, Georgetown University, Microsoft Research, National Institutes of Health y Office of Science and Technology Policy de la Casa Blanca (Kaggle).



(2013-16, MINECO, **HGU+HUB**)

[IxaMedTagger](#) entities



(2015-18, GV, **HGU+HUB**)



(2016-19, MINECO, **HGU+HUB**)



(2020-22, MINECO)



DTs expressed by doctors in MRs	ICD
Adenocarcinoma de prostata	185
Ca. prostata	185
CÁNCER DE PRÓSTATA	185
adenocarcinoma de pulmon estadio IV	162.9
CA pulmón estadio 4	162.9
ADENOCARCINOMA PÁNCREAS	157.9

Standard DTs from ICD-reference	ICD
NEOPLASIA MALIGNA PROSTATA	185
N. MALIGNA DE BRONQUIO Y PULMON SIN ESPECIFICAR	162.9
N. MALIGNA DE PANCREAS PARTE SIN ESPECIFICAR	157.9

DIAGNÓSTICO

Grp\_Enfermedad Causada\_por

-- EPIGASTRALGIA AUTOLIMITADA EN PROBABLE RELACION CON

Causada\_por Grp\_Medicamento

CON TRAMADOL

Grp\_Enfermedad

- LUMBOCIATGIA IZDA

- PREVIOS,

TRATAMIENTO

Grp\_Medicamento

Stop Zaldiar.

Grp\_Enfermedad Grp\_Medicamento

iene dolor además del Dolovoltaren puede to

Medicamento

Nolotil un comprimido cada ocho horas.

venía tomando

# Hizkuntzaren Azterketa eta Prozesamendua

## Language Analysis and Processing

(C) Joserra Martínez (Txerra) eta Sálcha Tovar



## <https://ixa.eus/master>

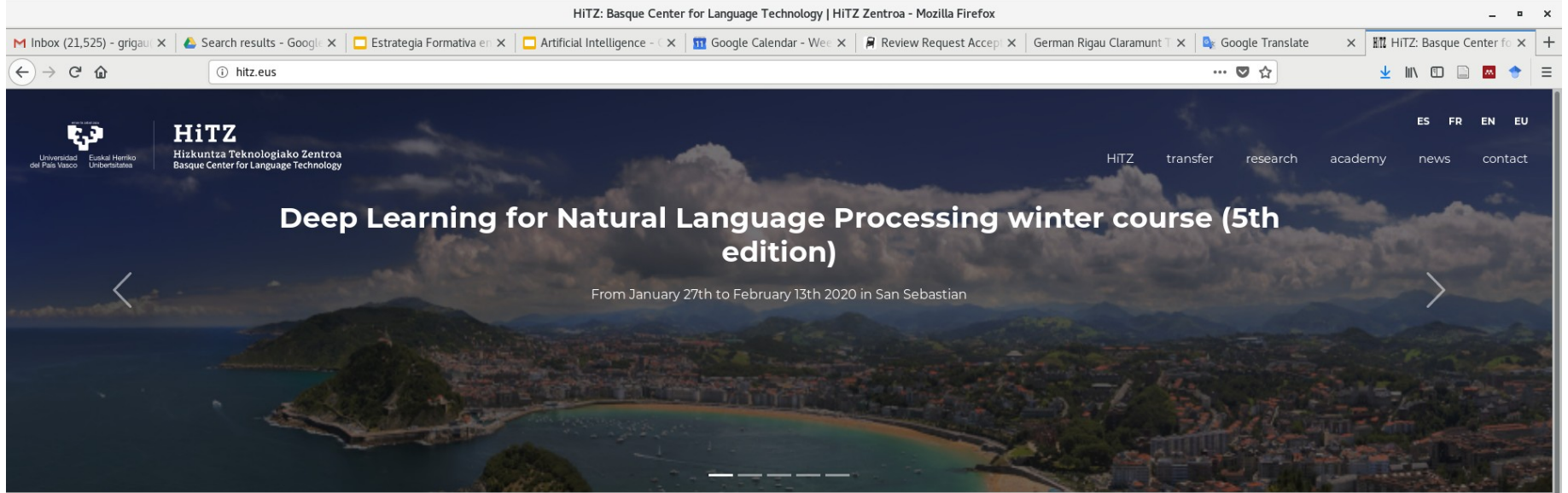
### Introduction

Advanced applications, based on the use of Language Analysis and Processing techniques, such as machine translation, automatic speech recognition and synthesis, dialogue systems, question answering, information extraction and web search systems are fast becoming an integrated part of our digital devices and daily lives. They are a key component in artificial intelligence-based solutions that provide people with natural ways to communicate with machines (e.g. Siri or Alexa), other people (e.g. machine translation in Skype), information repositories and much more.

These technologies require a mix of skills and approaches from computer science, linguistics, statistics, artificial intelligence, machine learning, logic... If you are a computer specialist, an engineer or a mathematician, you may be wondering about the basic computing tools working behind these devices or how you can integrate these applications within other applications or the web. If you are a philologist or linguist, you probably wonder how linguistic knowledge is used in them: lexicon, grammar and word-senses; morphological, syntactic and semantic analyses...







<http://hitz.eus/>

## HiTZ: Basque Center for Language Technology

HiTZ is a reference center on Language Technologies. Its aim is to promote research, training, technological transfer and innovation in this area. We are a multidisciplinary team: computer scientists, linguists and engineers.



**Information Extraction  
and Information  
Retrieval**



**Machine Translation**



**Text Analysis**



**Speech Synthesis and  
Speech Recognition**

# Artificial Intelligence

NLP and Deep Learning