Fourth Data science School

AI & Education (Nov. 6 - 10, 2023)

Leveraging NLP for Educational Empowerment

Dr. John Aoga, Data scientist & Programmers

MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
RÉPUBLIQUE DU BÉNIN

Mens Agitat Molem
UNIVERSITÉ D'ABOMEY-CALAVI

Depuis 1988

CEA-SMIA

BIOGRAPHY

Dr. John Aoga

Who am I ?

Doctor & Engineer in Science and Technology

Specialist in Data science & AI

Online Content author and Teacher

Co Founder of MIFY SARL start-up

Goals and Aspirations

Promote and develop AI 4 Africa In Africa

Promote and develop Education tools

Domains & Interests

Algorithms and Optimization

Data/Pattern Mining Approches and applications

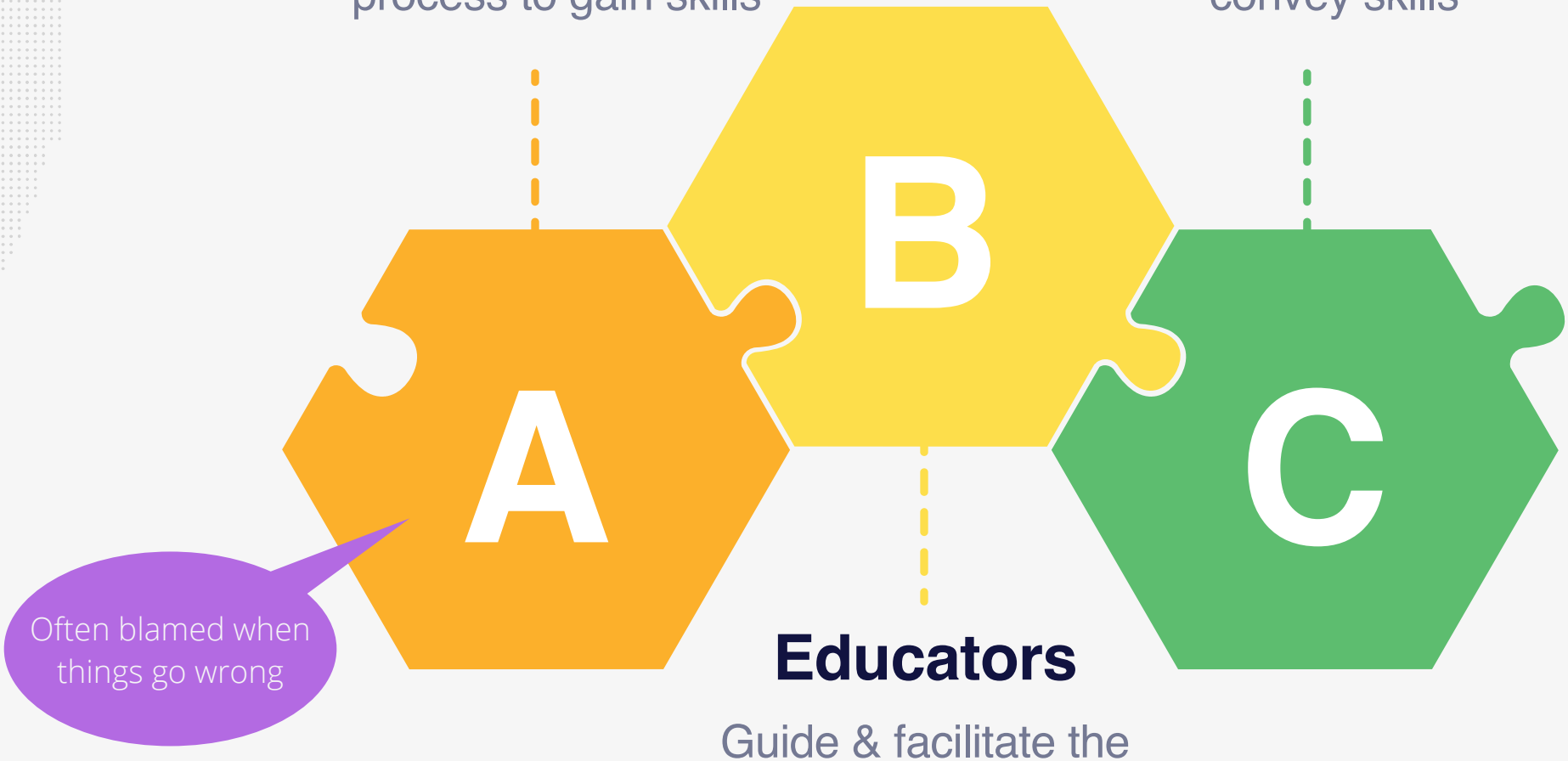
Deep Learning & NLP for local languages

Social Data Analysis

Scientific References

Criteria of Quality Education

Education is a (multifaceted) **process** that involves the **acquisition of knowledge,** skills, values, and attitudes through (structured) **learning experiences**.


Leaners
Focus on learning process to gain skills
Ressources
Educational material to convey skills
A
B
C
Often blamed when things go wrong
Educators
Guide & facilitate the learning process

SUSTAINABLE DEVELOPMENT GOALS

01

SDG principle

Sustainable Development Goals

02

No Poverty

Food sustainability, Outcome optimization, reduce inequalities

03

Quality Education

Online courses (+ create educational content)

04

Clean ressources

Clean water and sanitation,
Affordable and clean energy
Responsible consumption

05

Suitable Cities

Well design roads, transportation flows and optimization

06

Suitable Services

Digitalization of administration (Quick and efficient services)

07

Production

Optimization of production, Use of suitable Tools

08

Health

New ways to efficiently tackles healthcare problems

https://sdgs.un.org/goals

United Nations | Department of Economic and Social Affairs
Sustainable Development

Home | SDG Knowledge | Intergovernmental Processes | HLPF | SIDS | SDG Actions | Engage | News | About

Goals

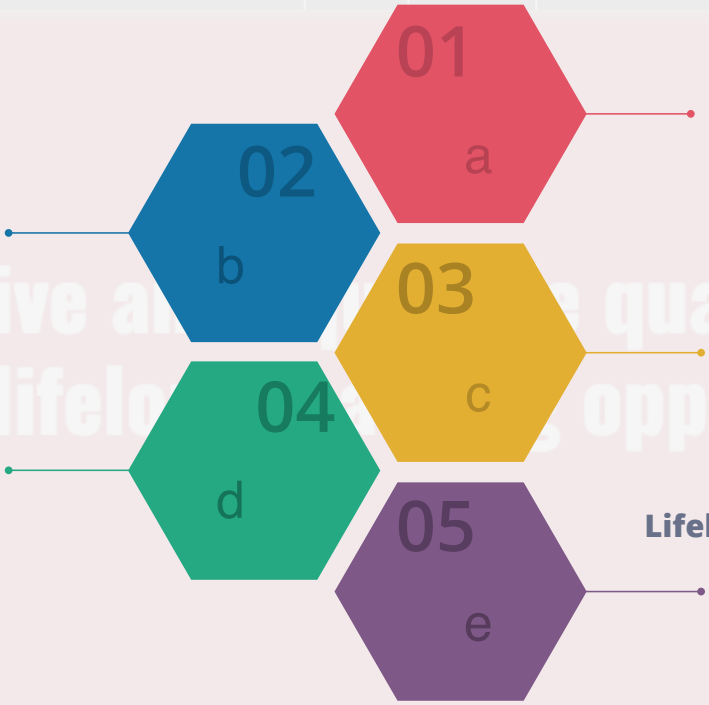
4

Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

← Previous Next →

https://sdgs.un.org/goals/goal4

AI & NLP can help here!



Accessible and Inclusive
To ALL, regardless of socioeconomic status, location, gender, or physical abilities

Relevant and Equitable
Tailored the needs of learners and society offering equal opportunities

High-Quality Teaching and Learning
Competent teachers and effective teaching methods (active learning and critical thinking)

Supportive Learning Environment
Physically safe and emotionally supportive, allowing students to learn without fear or discrimination

Lifelong Learning and Skills Development
development of essential life skills, such as problem-solving, communication, and digital literacy

https://sdgs.un.org/goals/goal4

EDUCATION'S CURRENT STATUS

The main ones

THE PANDEMIC

CAUSED LEARNING LOSSES IN 4 IN 5 OF 104 COUNTRIES STUDIED

Not Helping conditions

DESPITE SLOW PROGRESS,

THE WORLD IS FALLING FAR BEHIND IN ACHIEVING QUALITY EDUCATION

WITHOUT ADDITIONAL MEASURES, BY 2030:

84 MILLION
CHILDREN AND YOUTH WILL BE OUT OF SCHOOL

300 MILLION
STUDENTS WILL LACK BASIC NUMERACY/LITERACY SKILLS

ONLY 1 IN 6
COUNTRIES WILL ACHIEVE UNIVERSAL SECONDARY SCHOOL COMPLETION TARGET

LOW- AND LOWER-MIDDLE-INCOME COUNTRIES FACE A NEARLY

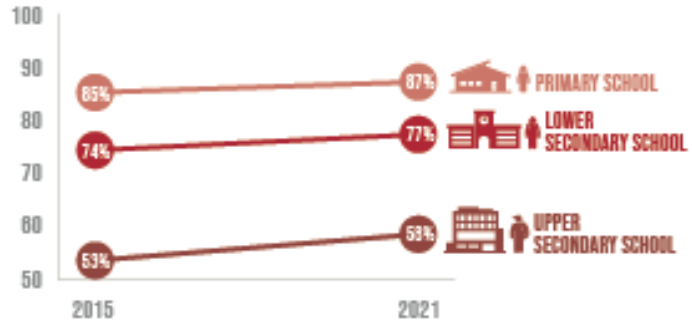
$100 BILLION

ANNUAL FINANCING GAP

TO REACH THEIR EDUCATION TARGETS

PRIMARY AND SECONDARY SCHOOL COMPLETION RATES ARE RISING, BUT THE PACE IS SLOW AND UNEVEN

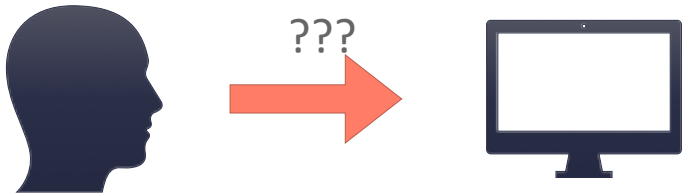
COMPLETION RATES



How AI and NLP can Help?

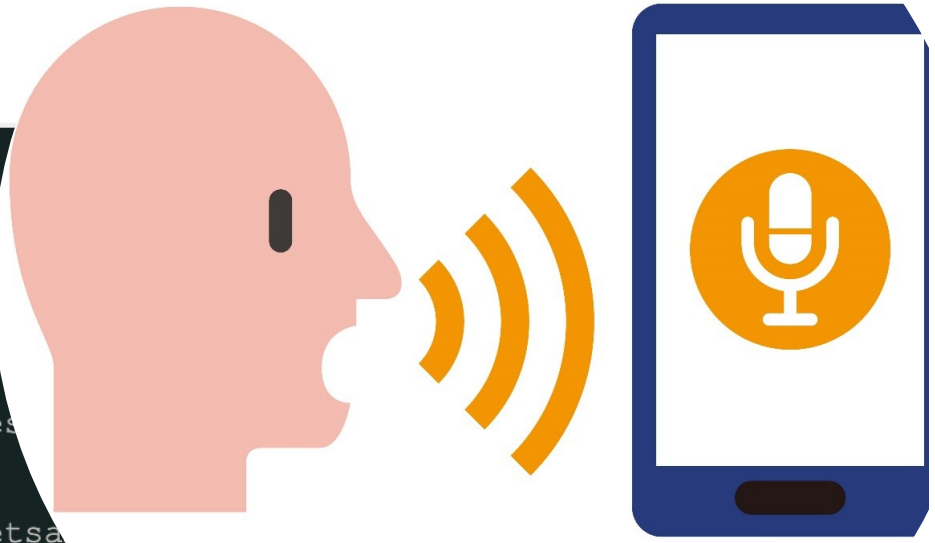
AI & Natural Language Processing

Humain & Computer communication



Before: Programming Language

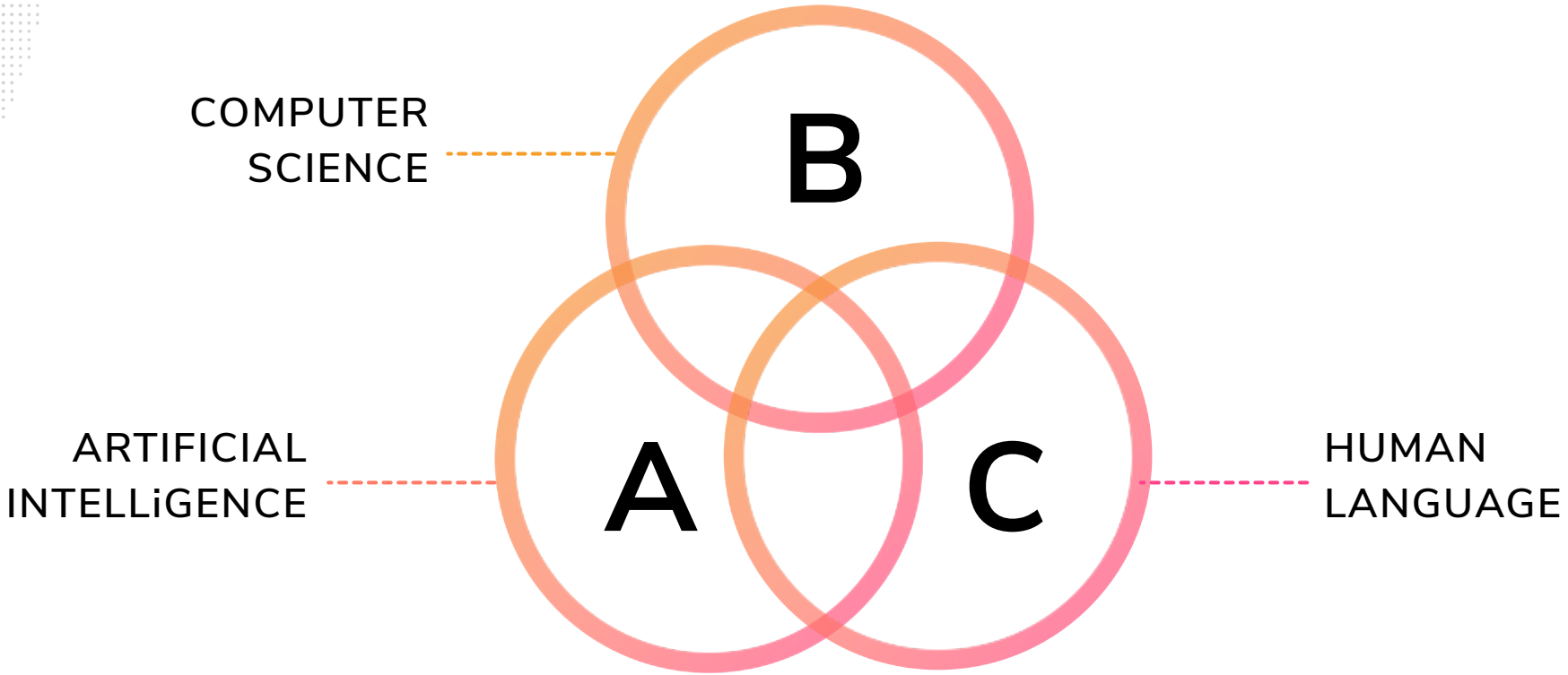
```
def play_audio_file(fname):
    """Simple callback function to play a wave file.

    :param str fname: wave file name
    :return: None
    """
    ding_wav = wave.open(fname, 'rb')
    ding_data = ding_wav.readframes(ding_wav.getnframes
    audio = pyaudio.PyAudio()
    stream_out = audio.open(
        format=audio.get_format_from_width(ding_wav.getsa
        channels=ding_wav.getnchannels(),
        rate=ding_wav.getframerate(), input=False, output=T
    stream_out.start_stream()
    stream_out.write(ding_data)
    time.sleep(0.2)
    stream_out.stop_stream()
    stream_out.close()
    audio.terminate()
```



After: Natural Language

What's NLP

Make computers «understand» and parse Natural(Human) Languages



Basic Apps of NLP

Two main components

@ NL Understanding

Mapping input to useful representation and Analyzing different of languages

1. Question and Answering
2. Sentiment analysis

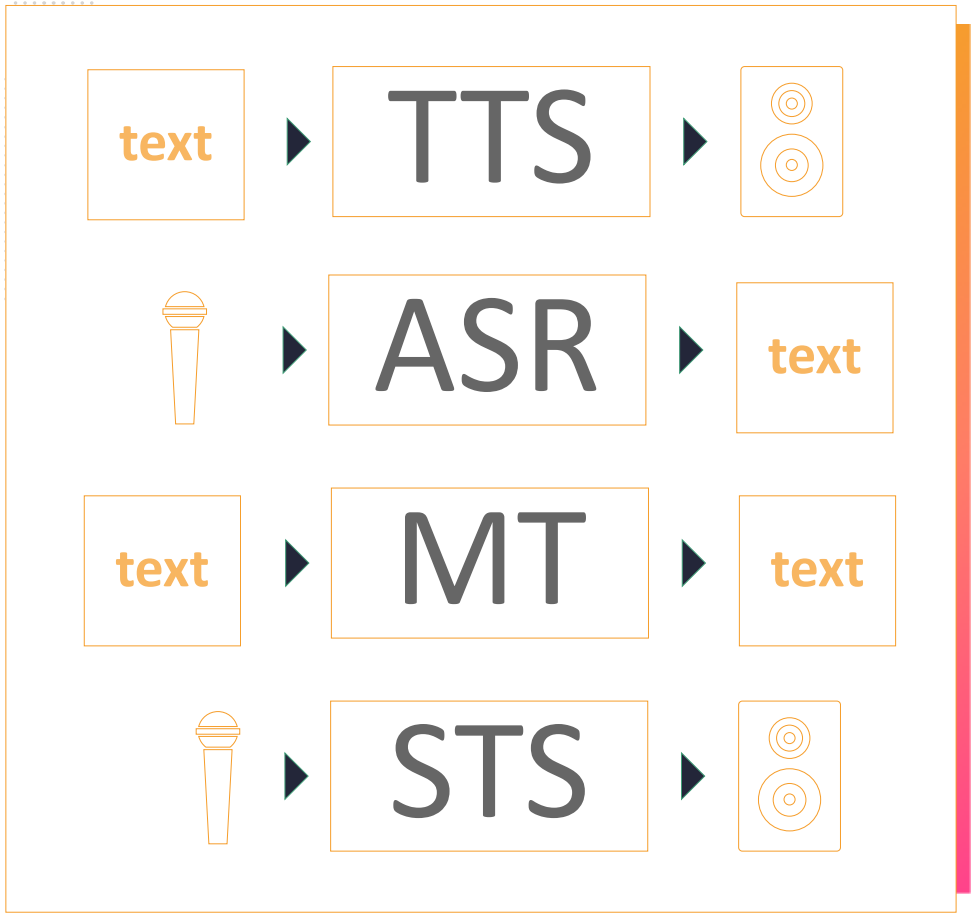
@ NL Generation

Produce meaningful phrases following a structure of a languages

1. Text Summarization
2. Text To Speech / Speech to Text
3. Machine Translation (Text & Speech)
4. Auto-completion / Story completion

Basic Apps of NLP

Four apps



NL Generation

Produce meaningful phrases following a structure of a languages

1. Text Summarization
2. Text To Speech / Speech to Text
3. Machine Translation (Text & Speech)
4. Auto-completion / Story completion

NLP vs Large Language Models

0 4 k e y p o i n t s h i g h l i g h t s h e r e

1 LLM is part of broad NLP field

2 LLMs are deep learning models trained to generate text and perform various NLP tasks

LLMs = advanced deep learning models (transformers) for massive language datasets

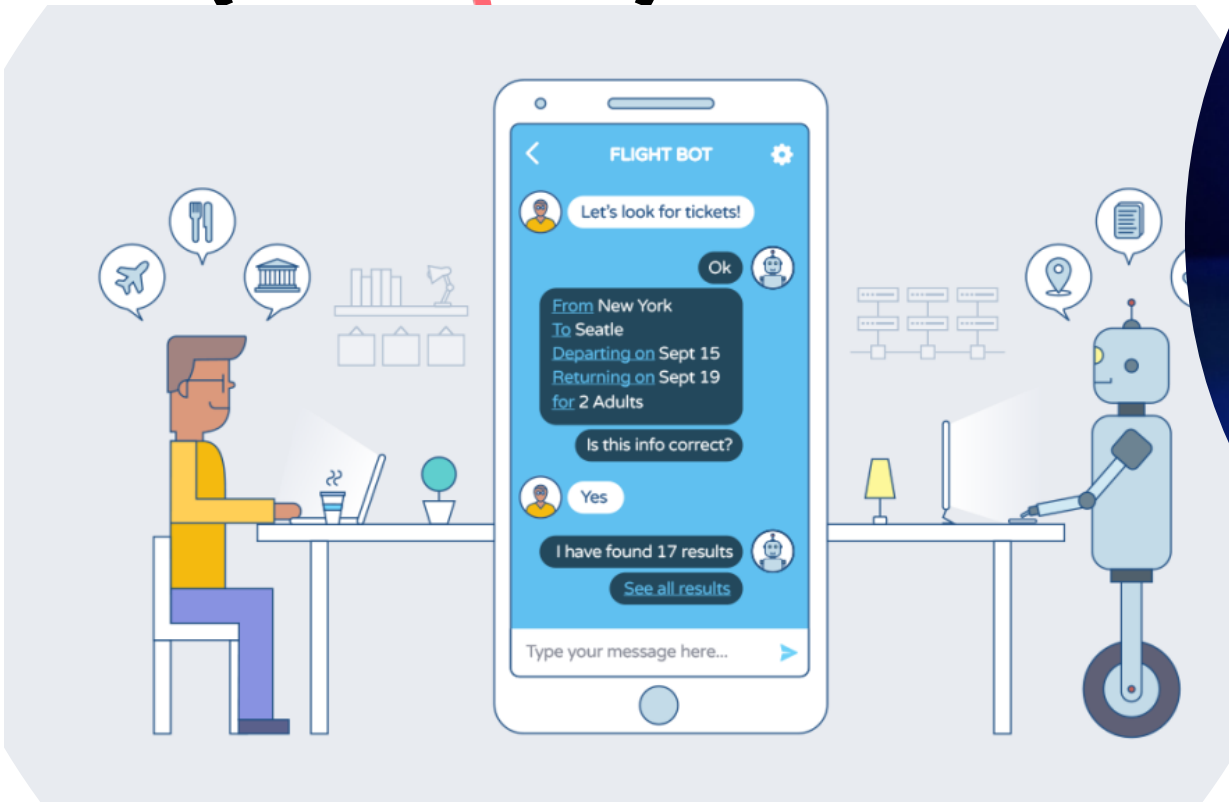
3 Text generation oriented

Design to mainly generate text

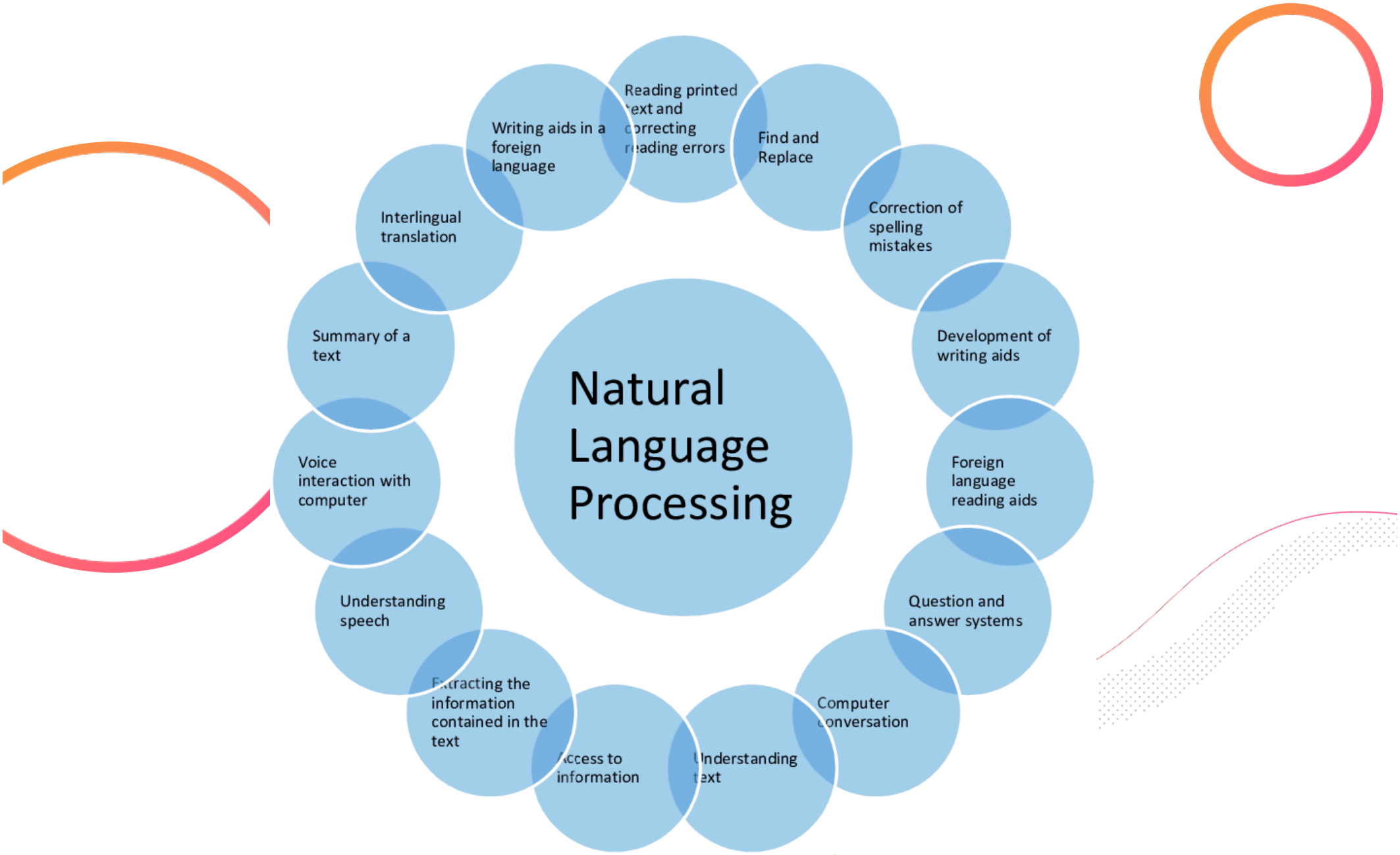
4 #Deep learning
#Transformers
#Attention mechanism
#Massive datasets
#Advanced Algorithms

NLP Applications

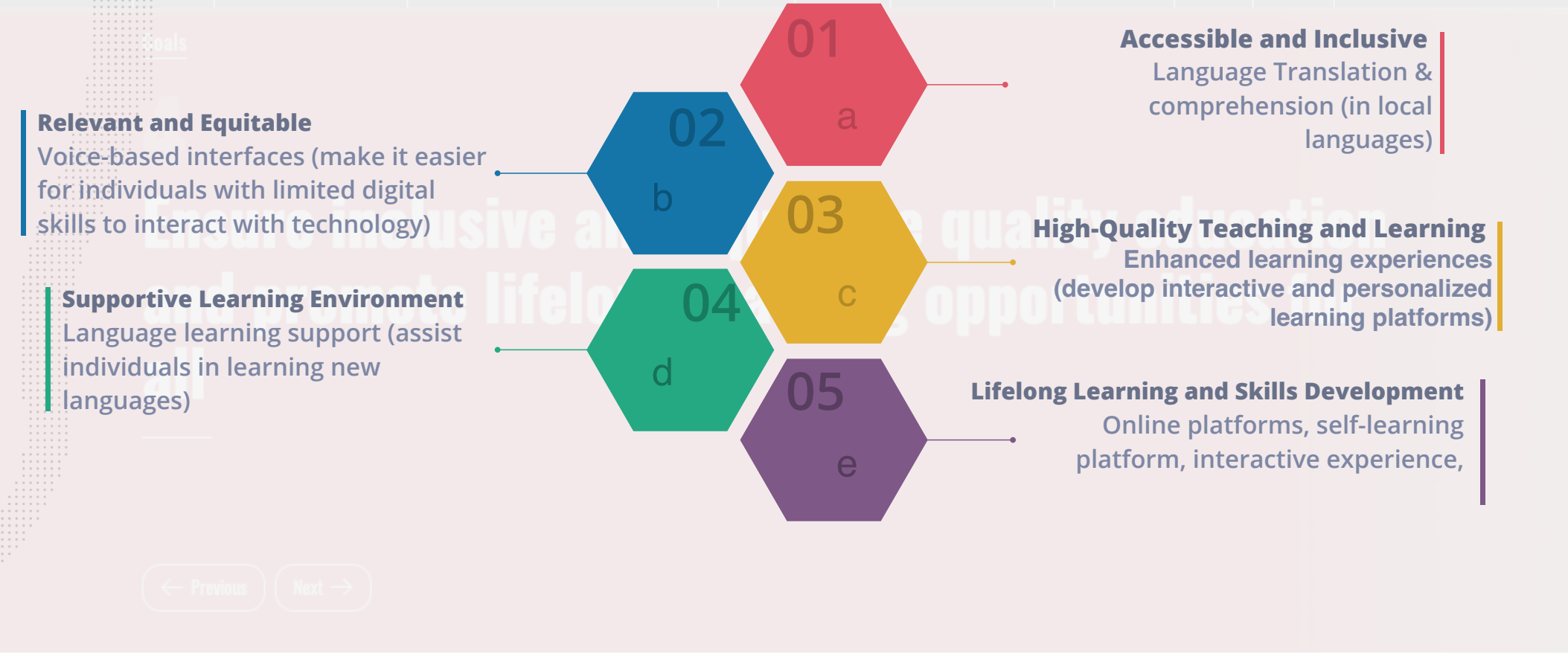
(Chat)bots



Voice Assistant


Natural Language Processing
Reading printed text and correcting reading errors
Find and Replace
Correction of spelling mistakes
Development of writing aids
Foreign language reading aids
Question and answer systems
Computer conversation
Understanding text
Access to information
Extracting the information contained in the text
Understanding speech
Voice interaction with computer
Summary of a text
Interlingual translation
Writing aids in a foreign language

Many more: Be creative!



https://sdgs.un.org/goals/goal4

NLP apps for each components

Learners

EXAMPLE: AI & NLP SUPPORTING LEARNING

- Customized learning (recommends specific lessons, exercises, or resources tailored to each student's needs, optimizing their learning experience)
- Automated Grading and Feedback (immediate feedback allows students to identify their mistakes and improve their performance)
- Chatbots for Tutoring (students can ask questions, seek explanations, or request assistance with homework…providing instant support)
- Content Recommendation and Search (suggest relevant educational content, such as articles, videos, or online courses, to help students explore topics aligned with their interests and educational goals)

Educators

EXAMPLE: AI & NLP SUPPORTING (ONLINE) TEACHING

- Provide Tools to share work (virtual classroom,)
- Produce content (remove audio/video noise, correct video, add caption in other language, video summary, translate into another language)
- Build Exercises and Evaluate Students (Auto grading, promote interactions)
- Knowledge acquisition diagnostics and student profiling
- Breaking barriers (easy-to-use platforms, drive motivation in using platform, …)

The System

EXAMPLE: AI SUPPORTING ONLINE PRODUCTION

- Content Creation and Generation
- Adaptive Learning Resources
- Textbook Summarization and Annotation
- Language Translation and Accessibility
- Sentiment Analysis in Educational Content
- Plagiarism Detection
- Content Enhancement for Special Needs Students

text-to-speech applications convert text-based materials into audio format, helping visually impaired students.

Strategy

RESEARCH STRATEGIES

Two ways

PARADIGMATIC

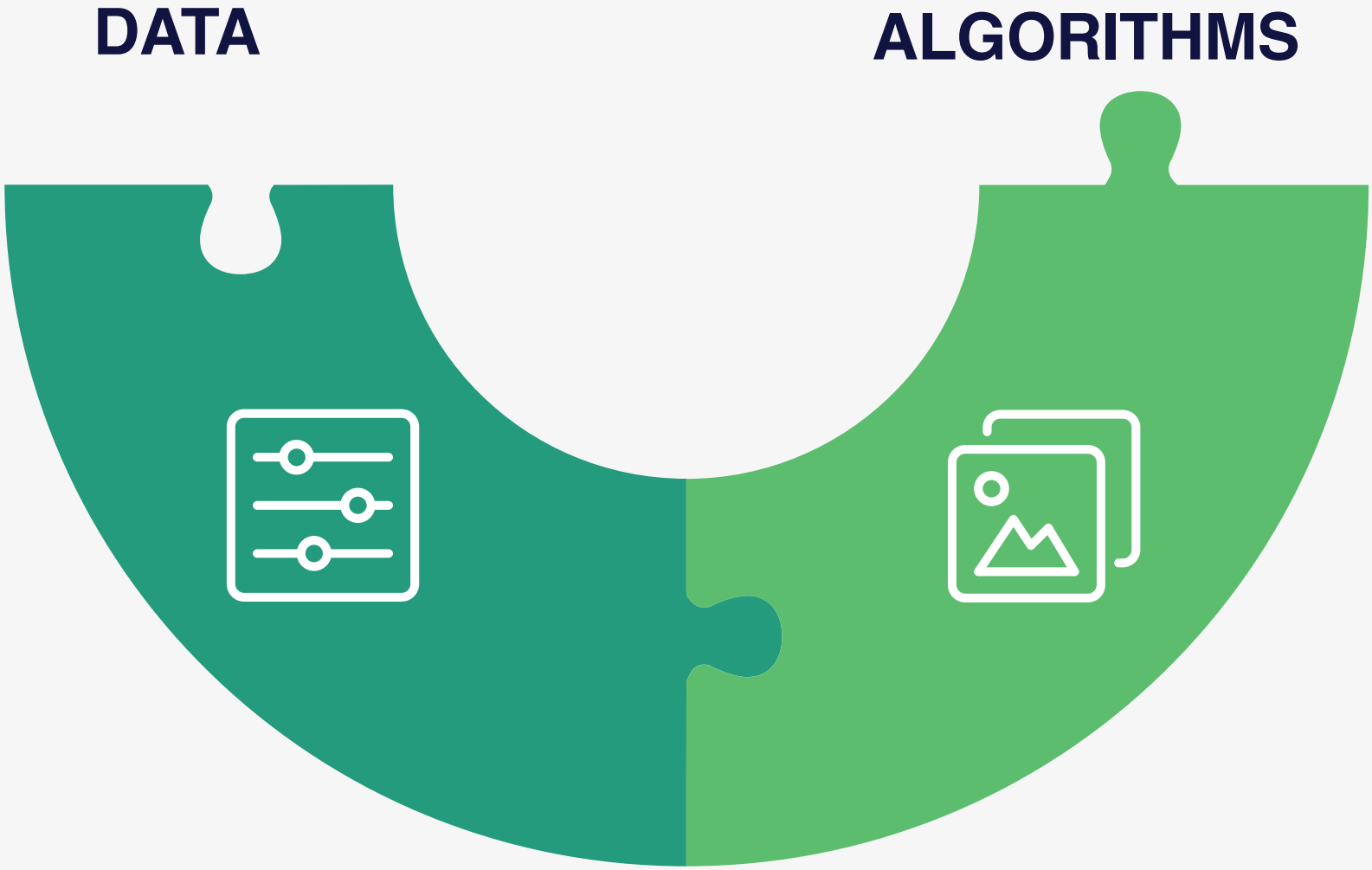
Existing research on development issues => Uncover the specific contributions

CRITICAL

Investigate the normative implications entailed by the AI disruption of the field of development

GLOBAL STRATEGIES

Two ways in one



Andrew Ng, Stanford University

GALLUP DATA EXAMPLE

Two components

- ☑ Feasibility of applying to development issues
  - Existing of appropriate data
  - Challenge of scholars and policy makers to tailor AI to needs and priorities
- ☑ Performance gaps
  - AI promises vs AI delivers
  - (COVID-19 as accelerator) Easy-to-use platforms

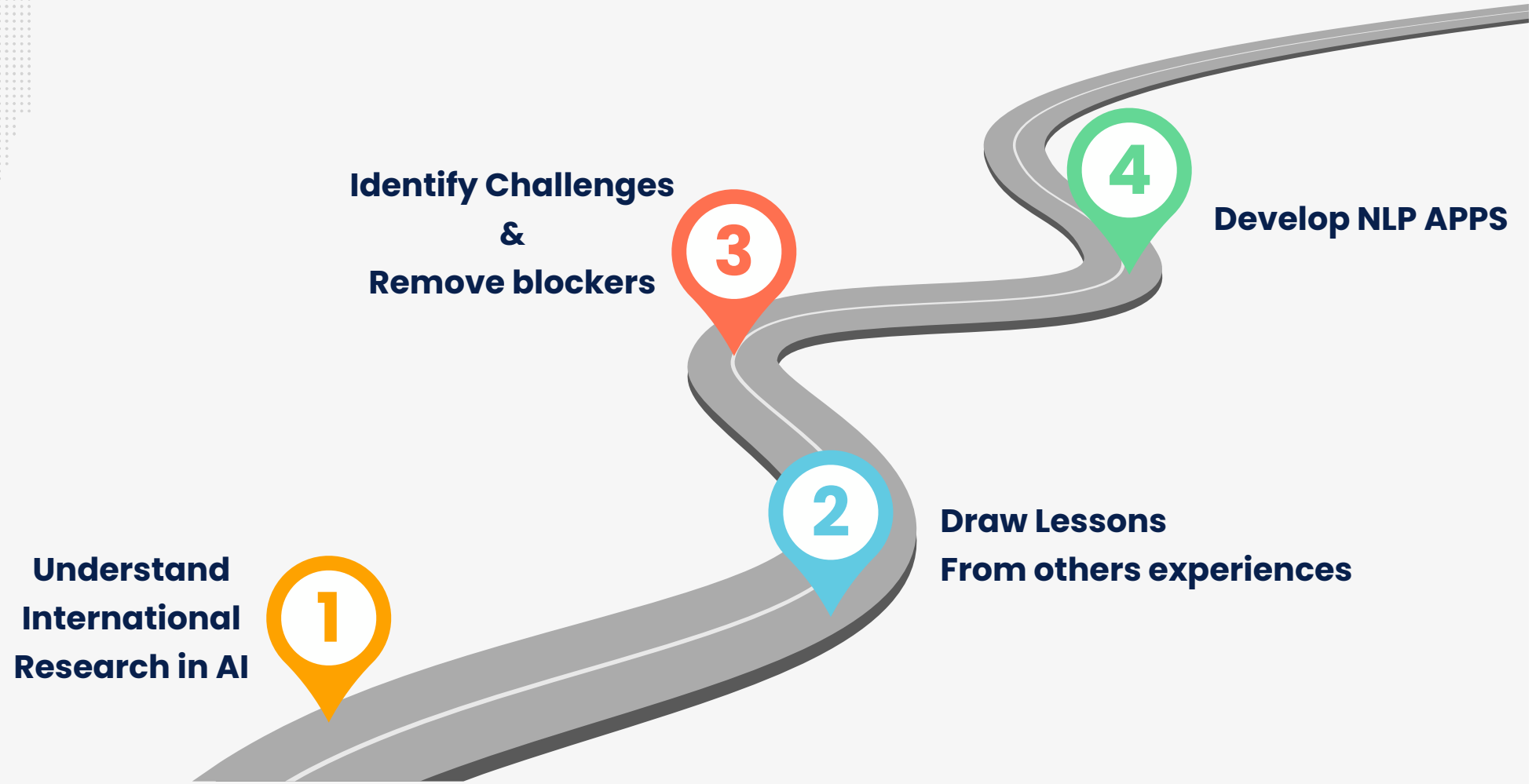
Andrew Ng, Stanford University

GLOBAL RESEARCH STRATEGIES

- Invest in Education (For AI & in AI)
  - Use NLP tools to convert existing educational content and other in our languages
  - Use NLP and other AI approaches to make online education effective (auto-grading, peer evaluation, community handling)
- Promote Cross-Fields Research
  - Weather, Economy & AI
  - Food Sustainability & AI

Ethics Considerations!

Takeaways


Understand International Research in AI
1
Draw Lessons From others experiences
2
Identify Challenges & Remove blockers
3
Develop NLP APPS
4

Fourth Data science School

AI & Education (Nov. 6 - 10, 2023)

Leveraging NLP for Educational Empowerment

Dr. John Aoga, Data scientist & Programmers

MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
RÉPUBLIQUE DU BÉNIN

Mens Agitat Molem
UNIVERSITÉ D'ABOMEY-CALAVI

Depuis 1988

CEA-SMIA

Techs Adventure

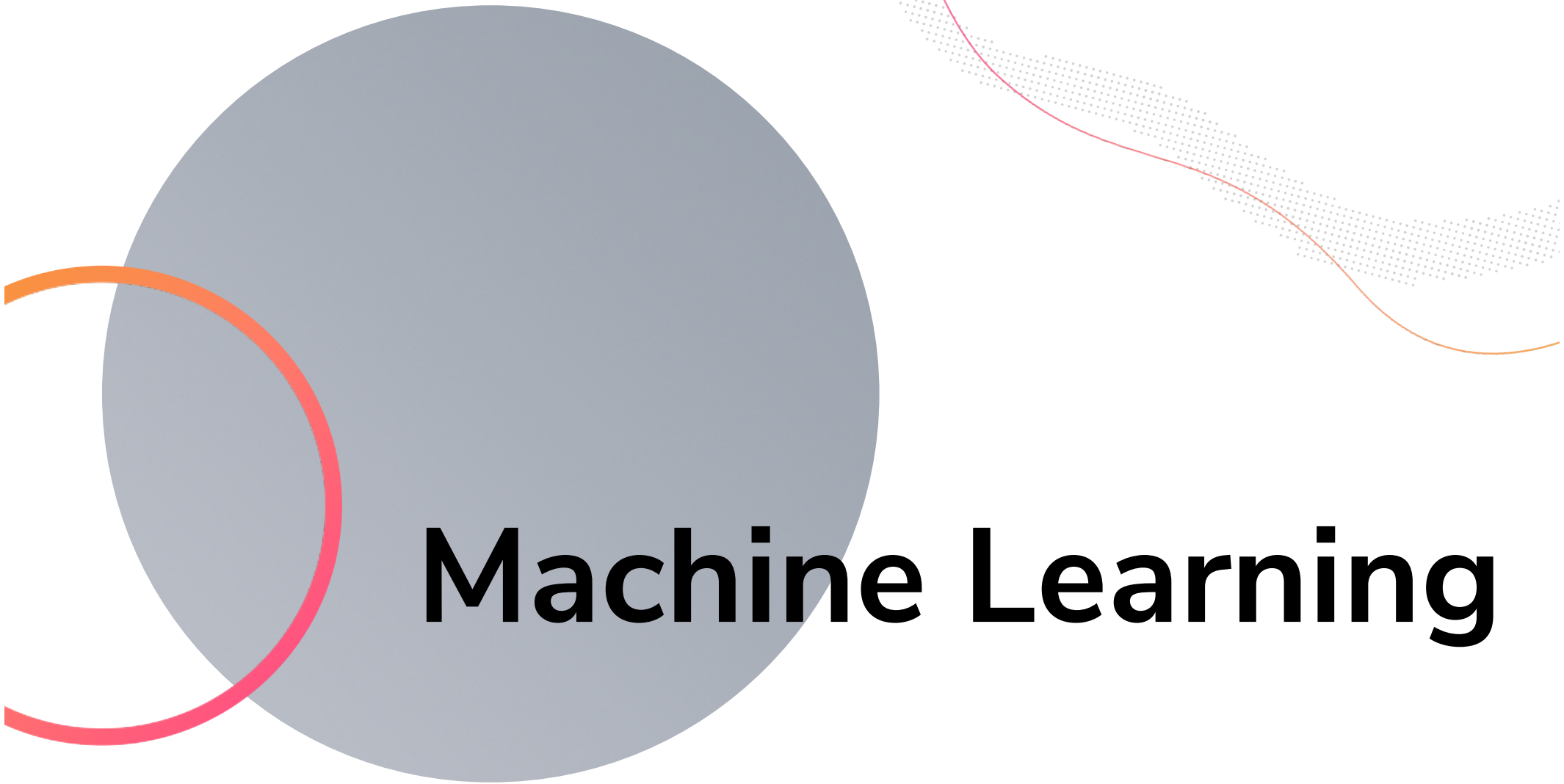
MIFY,
AS A PIONEER ?

ASSISTANT VOCAL CICA

PERMET DE LANCER DES COMMANDES SUR SON PC OU TÉLÉPHONE

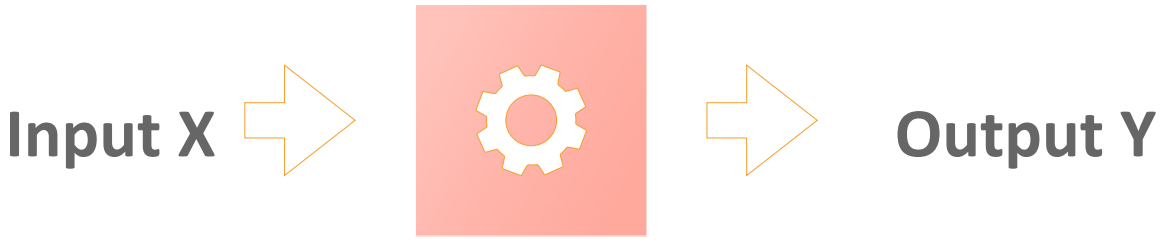
ASSISTANT VOCAL CICA

PERMET DE LANCER DES COMMANDES SUR SON PC OU TÉLÉPHONE

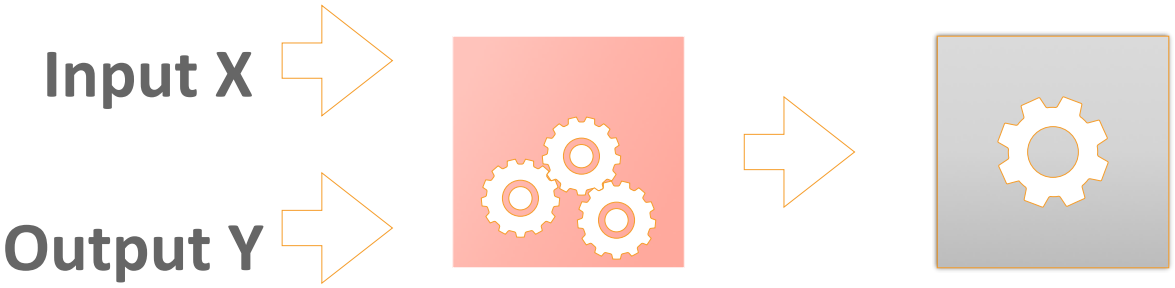

Machine Learning

ML principles

Basics



1, 2, 3, 4,… —> x^2 —> 1, 4, 9, 16, …

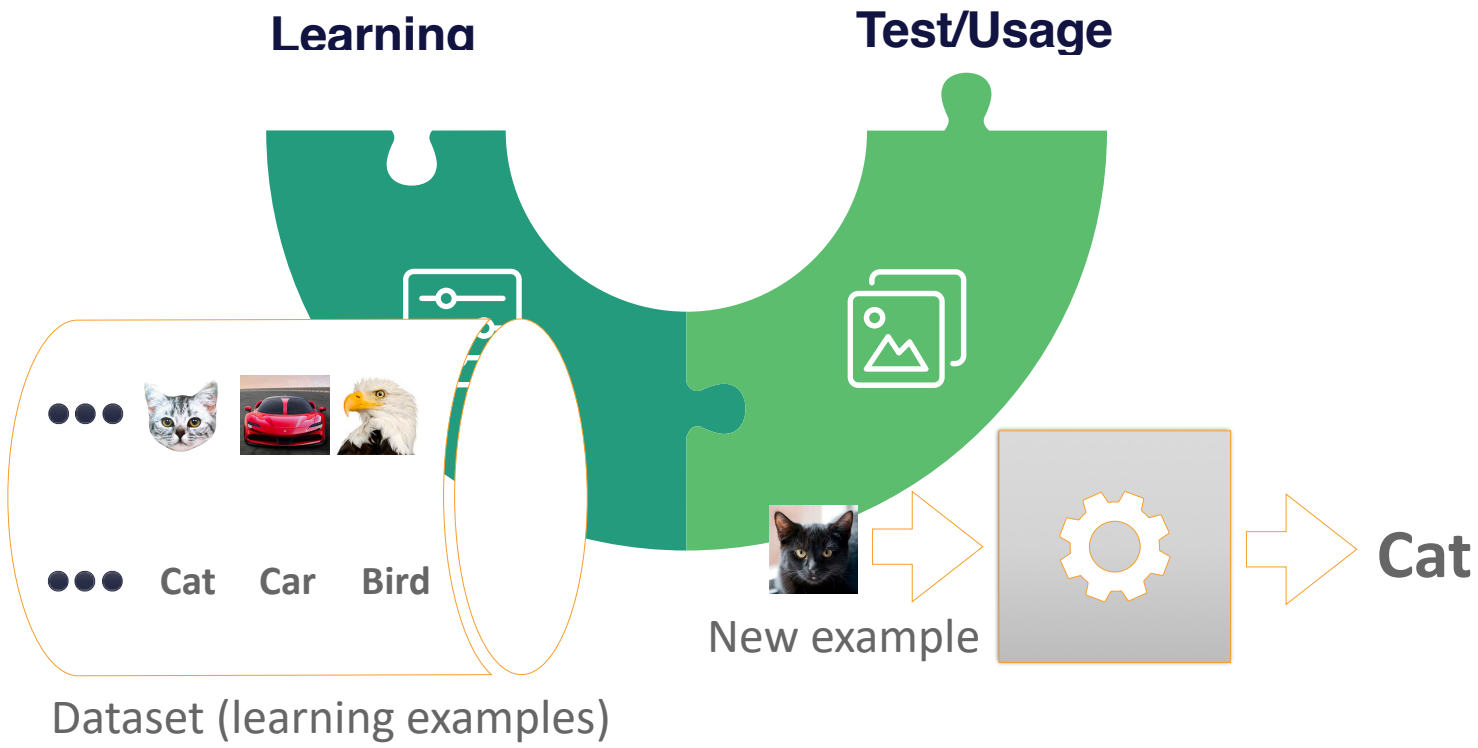


1, 2, 3, 4,…
&
1, 4, 9, 16, …

—> ? —> x^2

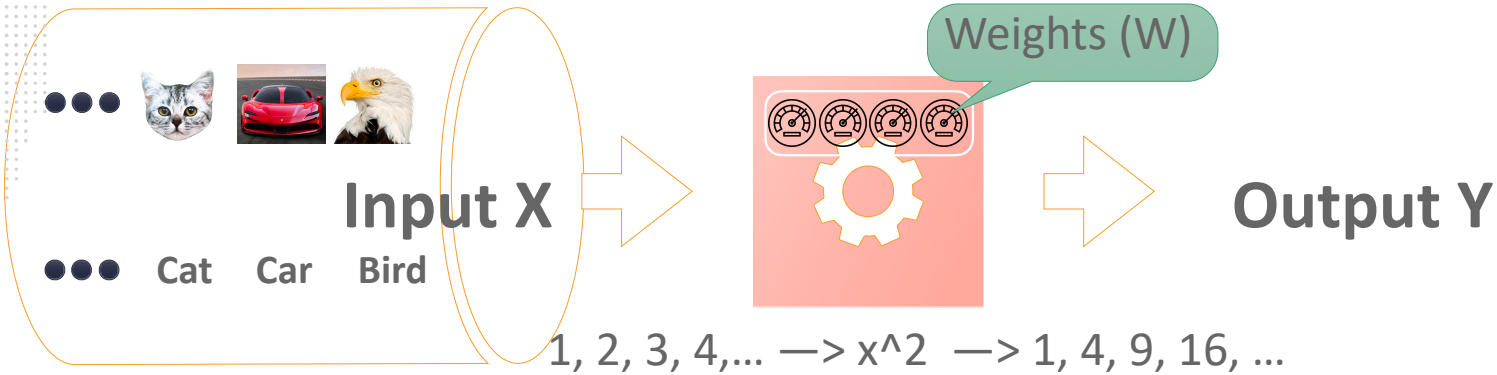
ML Phases

Learning & tests

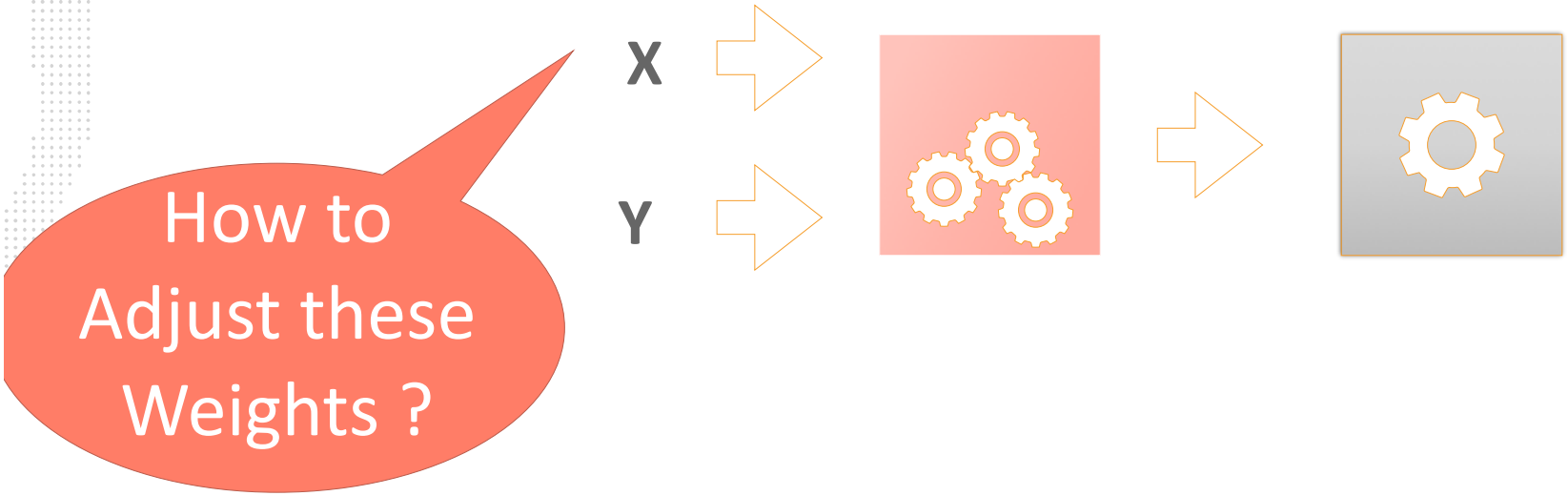


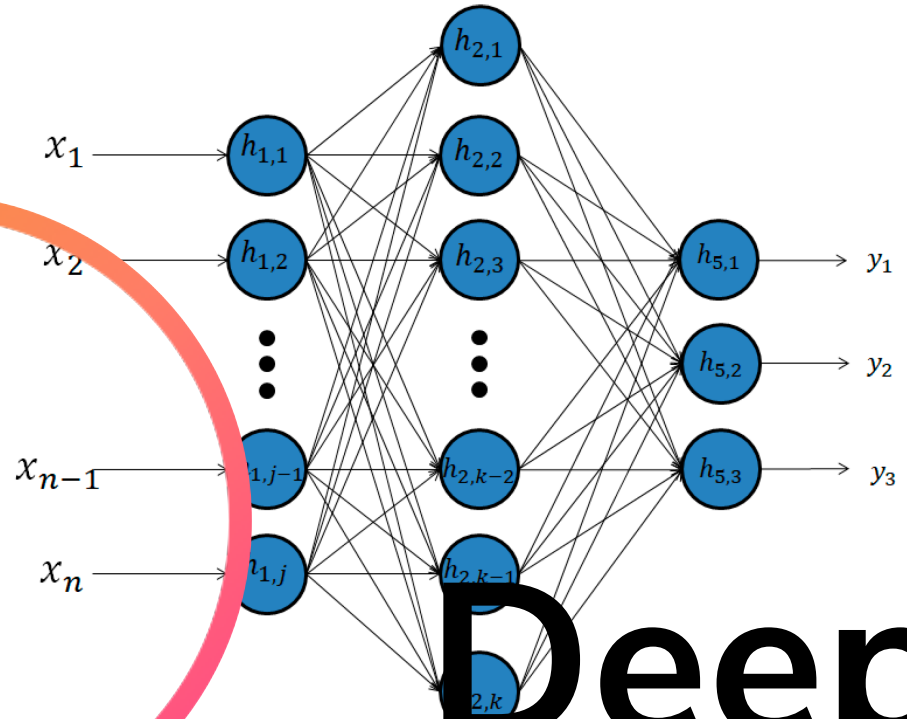
ML Phases

Learning & tests



1, 2, 3, 4,… —> x^2 —> 1, 4, 9, 16, …

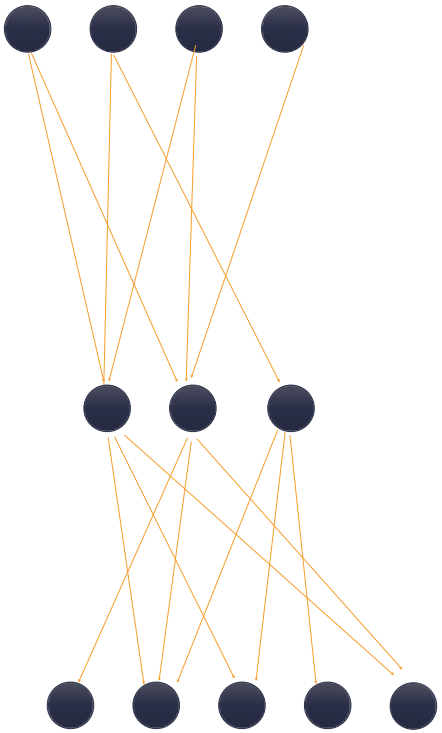




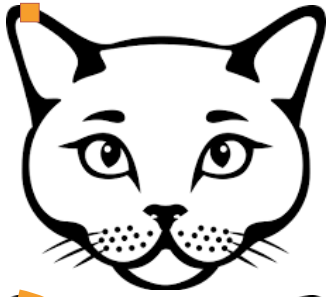
Deep Learning

Convolutional Neural Network

Cnn

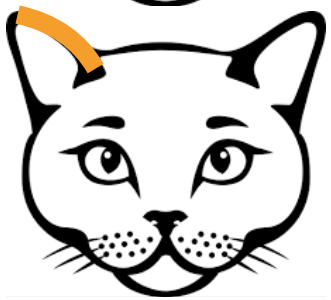


Pixels



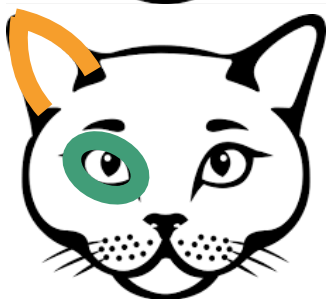
Pixels differences

Vectors



Combines pixels (vectors)

Shapes

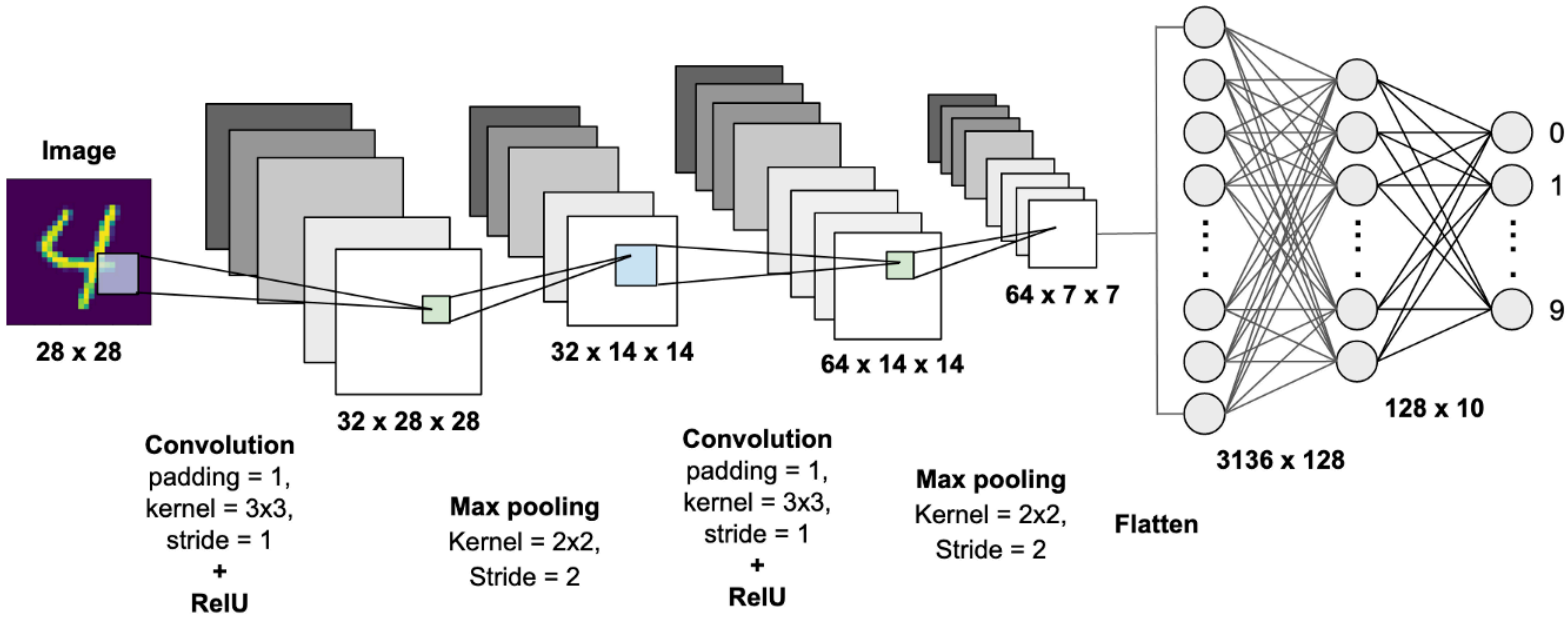


Combine vectors (shapes)

Combine Shape to identify the **CAT**

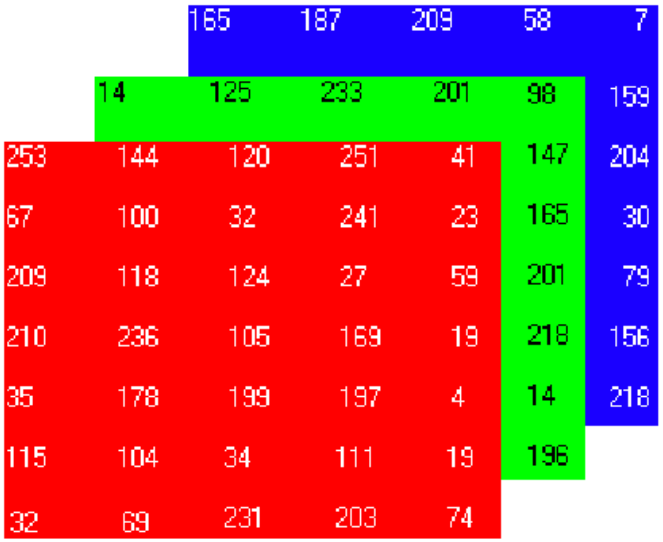
Convolutional Neural Network

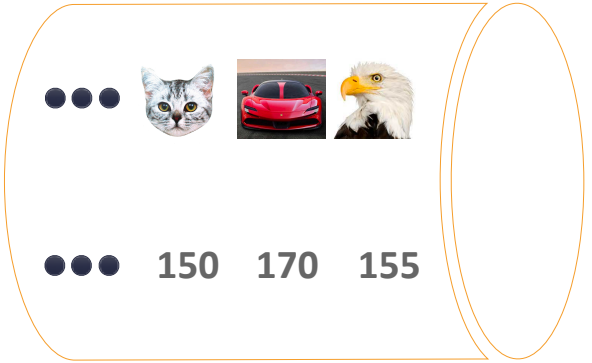
Cnn



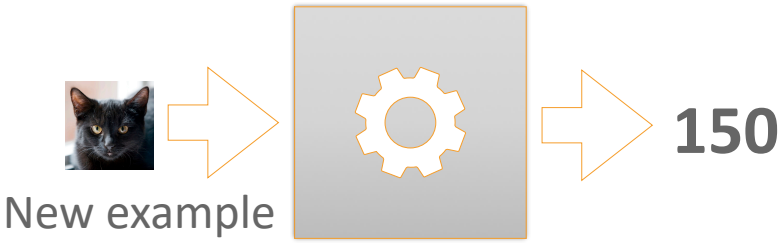
Supports only numbers

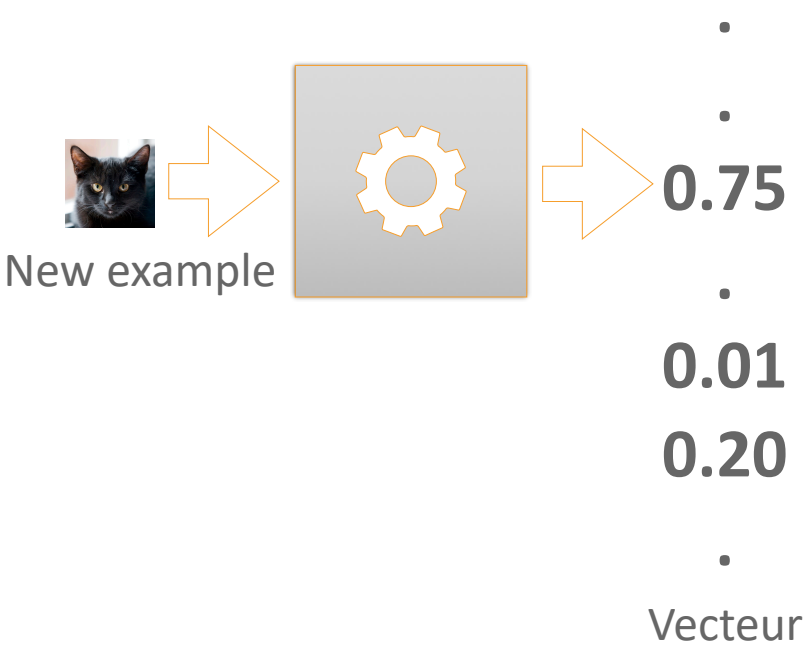
Solve number's Problems

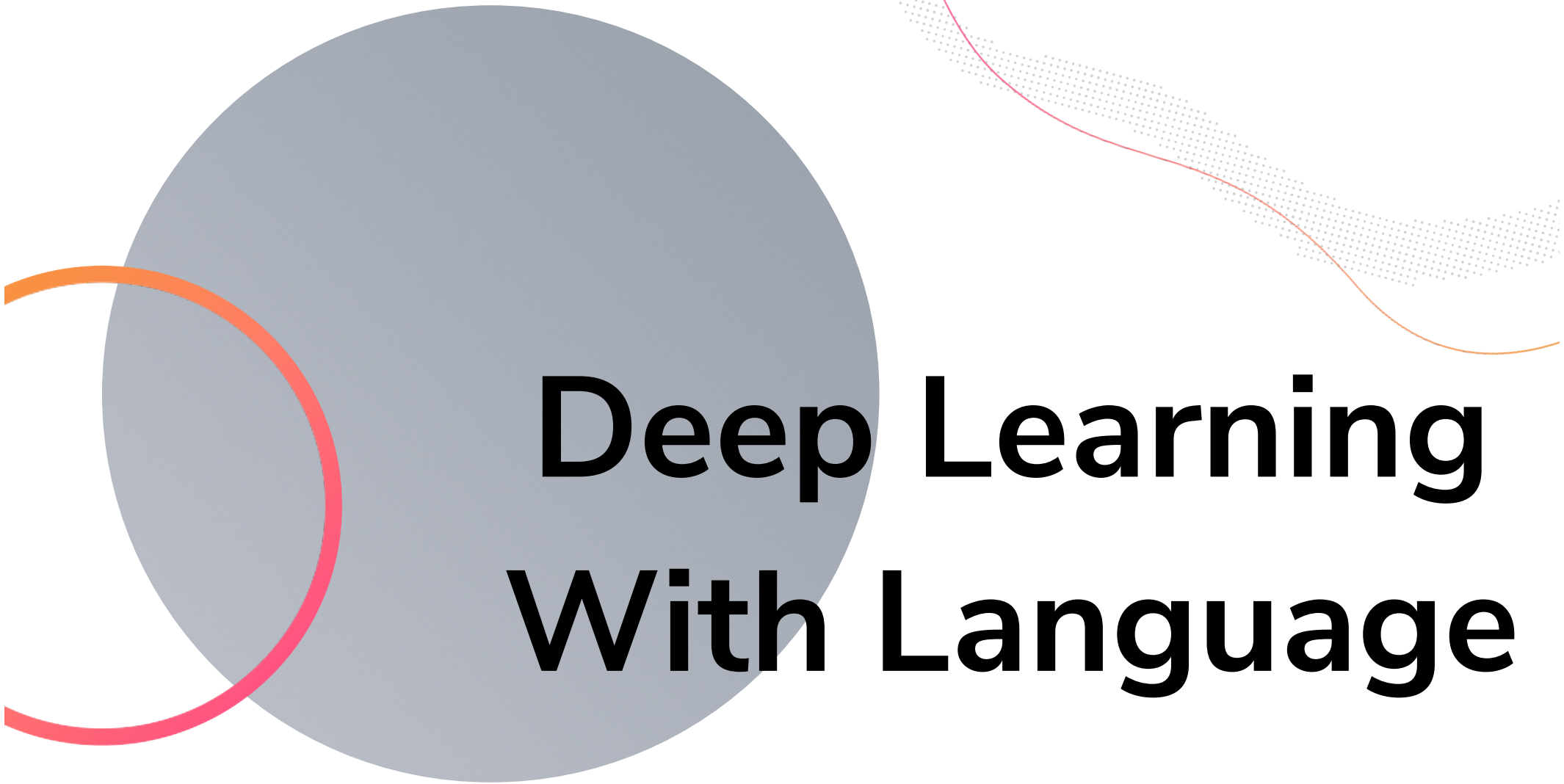




Solve number's Problems






Deep Learning
With Language

Language Processing Pain Points

04 pain points highlights here

1 **Text -> numbers**

2 **Text with different sizes**

Image with fix size at the begin, if not resize

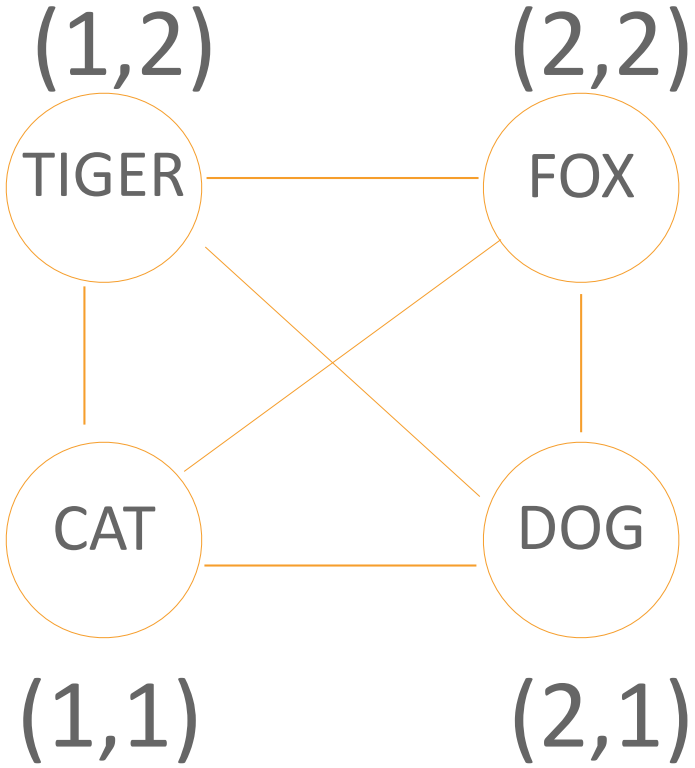
3 **Order is really important**

Difficult to split the sentence without loosing information

4 **Meaning**

Emotion detection in a sentence

Text to numbers



In Real, more than 300 dimensions

Word2vec
GloVe

Glove - Google NLP

GloVe: Global Vectors for Word Representation

Jeffrey Pennington, Richard Socher, Christopher D. Manning

Introduction

GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.

Getting started (Code download)

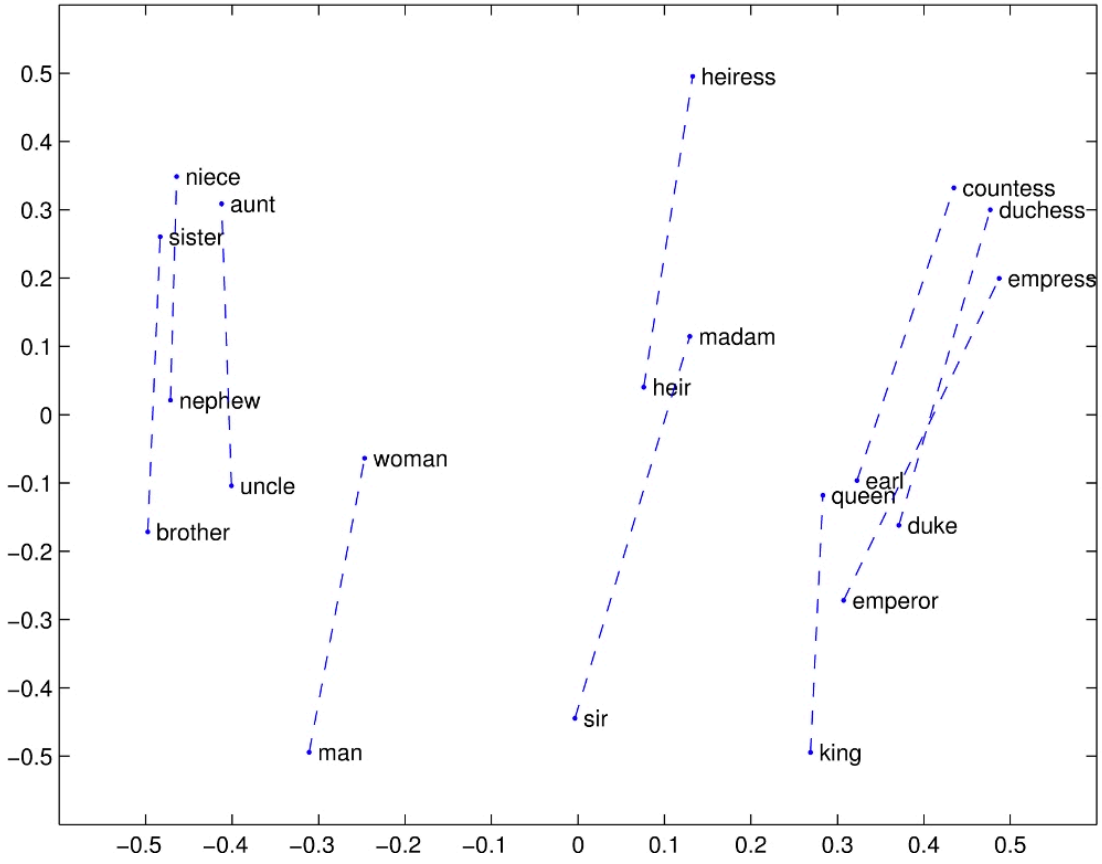
- Download the latest latest code (licensed under the Apache License, Version 2.0).
  Look for "Clone or download"
- Unpack the files: unzip master.zip
- Compile the source: cd GloVe-master && make
- Run the demo script: ./demo.sh
- Consult the included README for further usage details, or ask a question

Download pre-trained word vectors

- Pre-trained word vectors. This data is made available under the Public Domain Dedication and License v1.0 whose full text can be found at: http://www.opendatacommons.org/licenses/pddl/1.0/.
  - Wikipedia 2014 + Gigaword 5 (6B tokens, 400K vocab, uncased, 50d, 100d, 200d, & 300d vectors, 822 MB download): glove.6B.zip
  - Common Crawl (42B tokens, 1.9M vocab, uncased, 300d vectors, 1.75 GB download): glove.42B.300d.zip
  - Common Crawl (840B tokens, 2.2M vocab, cased, 300d vectors, 2.03 GB download): glove.840B.300d.zip
  - Twitter (2B tweets, 27B tokens, 1.2M vocab, uncased, 25d, 50d, 100d, & 200d vectors, 1.42 GB download): glove.twitter.27B.zip
- Ruby script for preprocessing Twitter data

https://nlp.stanford.edu/projects/glove/

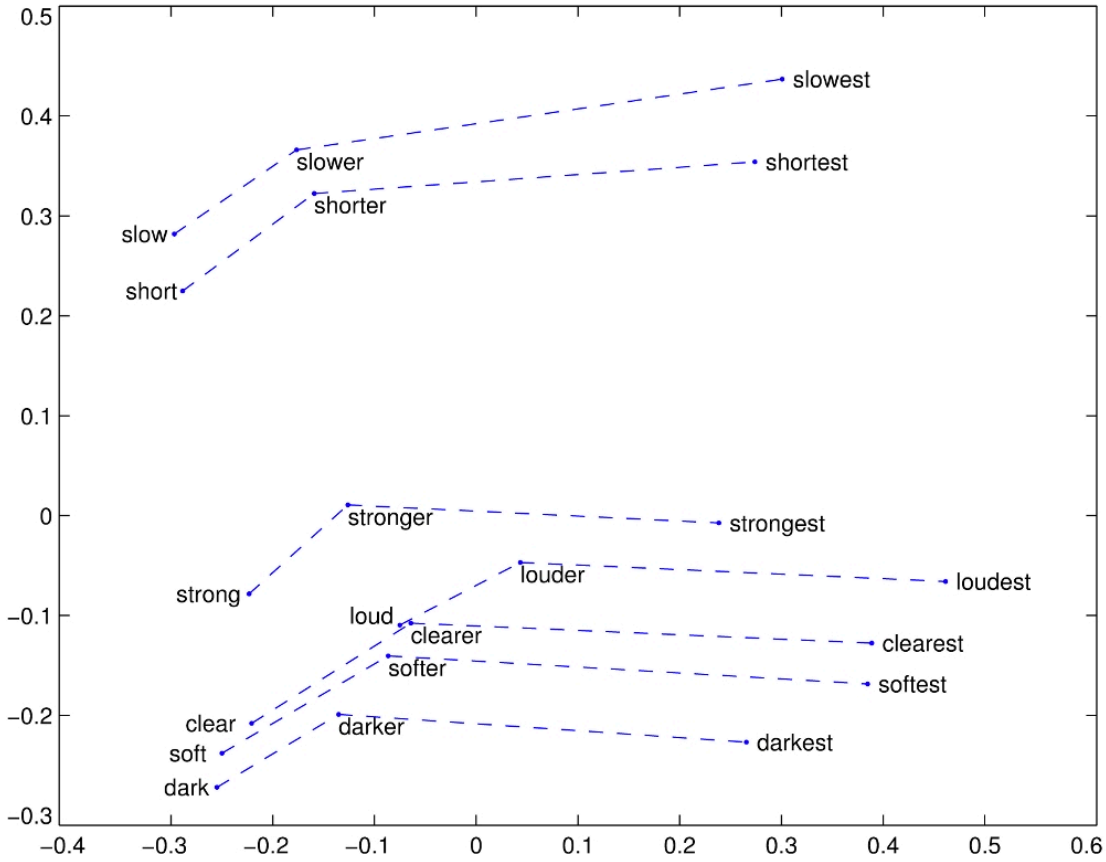
Glove - Google



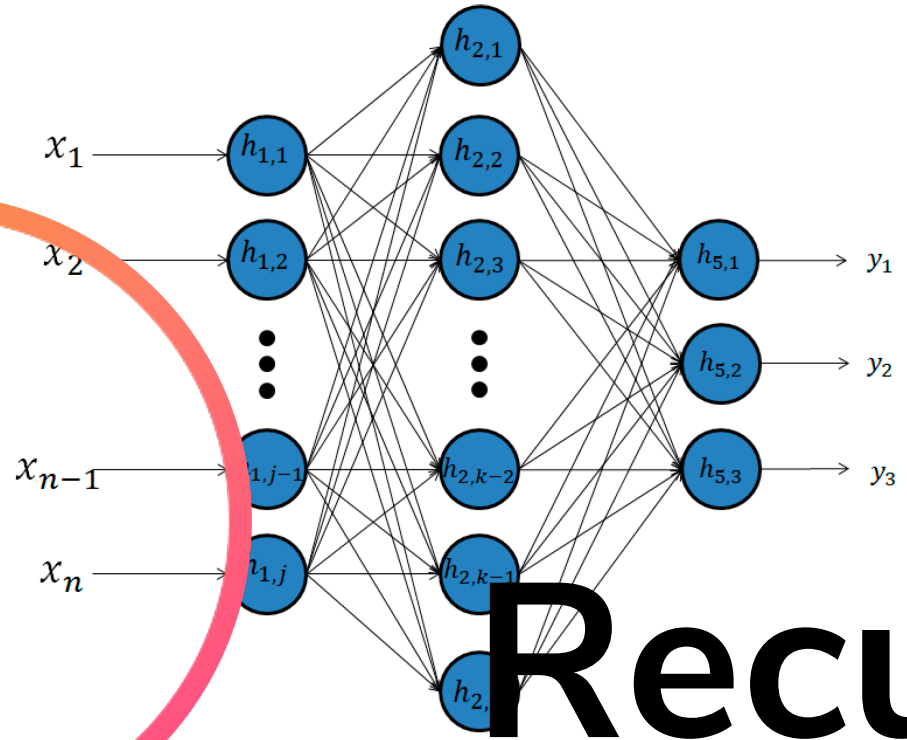
https://nlp.stanford.edu/projects/glove/

Glove - Google

N L P



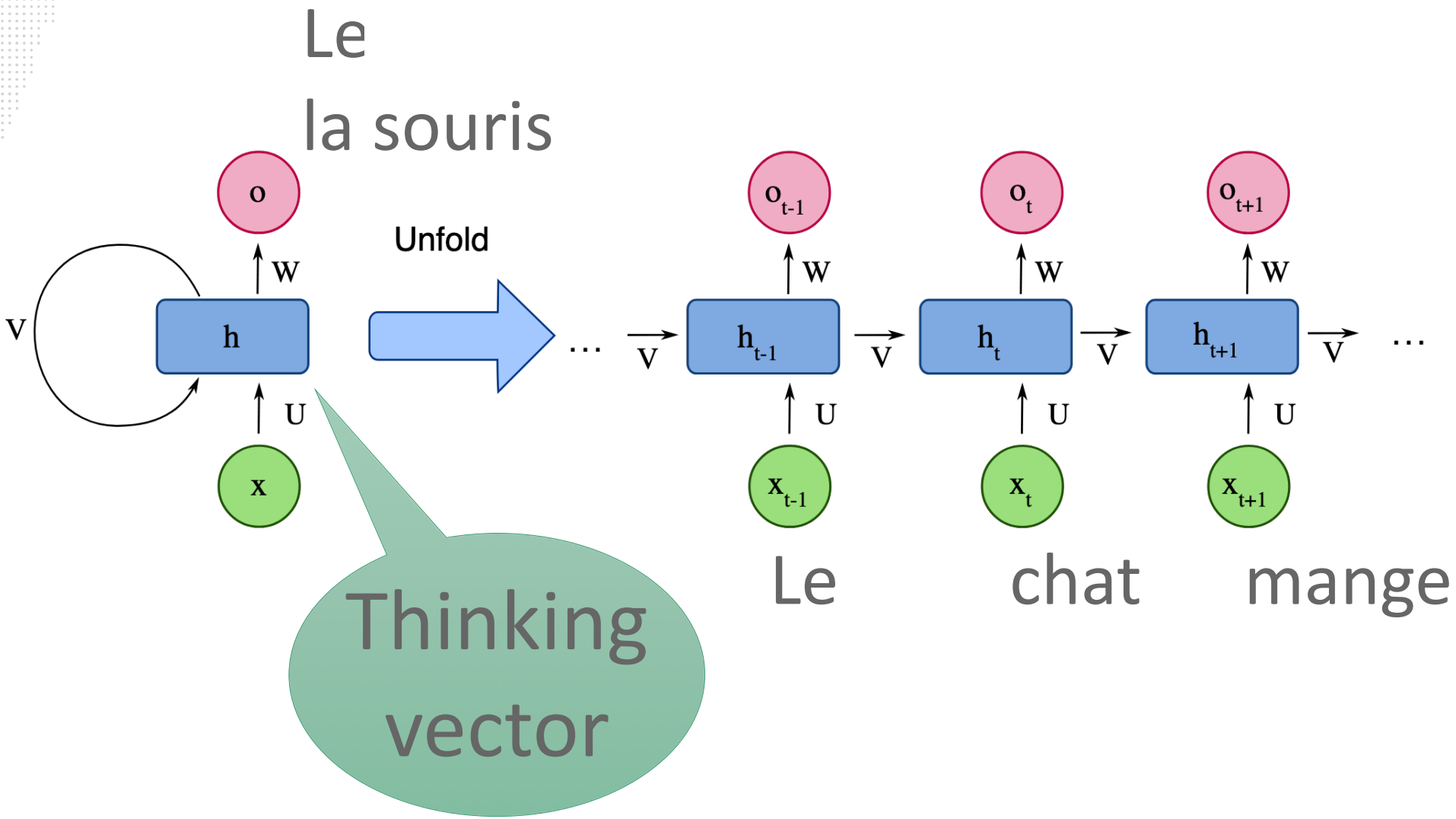
https://nlp.stanford.edu/projects/glove/

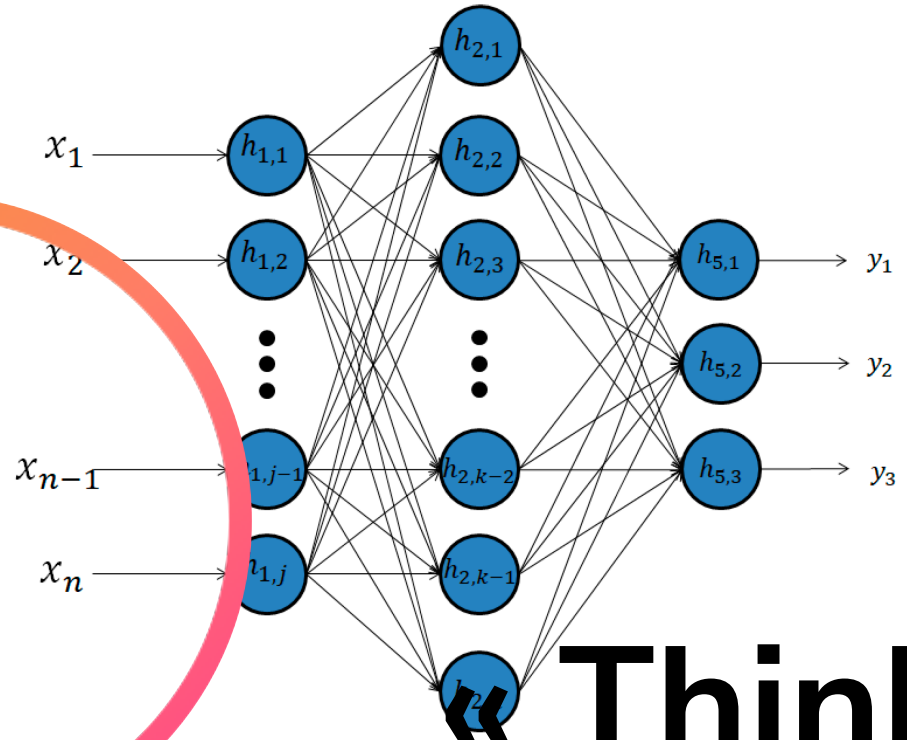


Recurrent NN

Recurrent Neural Network

R N N

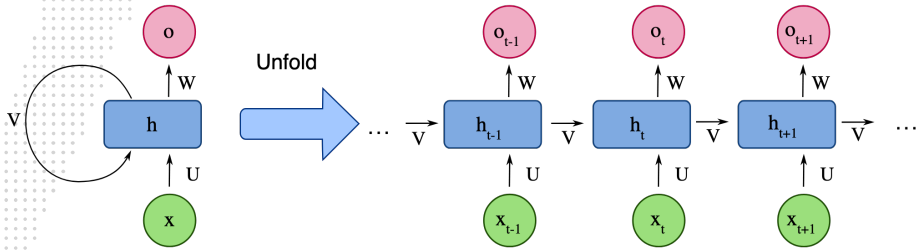




« Thinking » vector?

Recurrent Neural Network

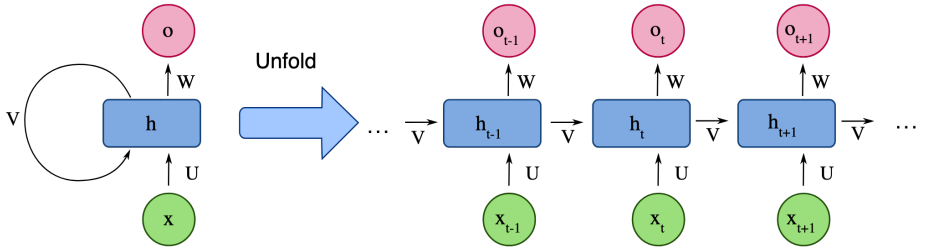
R N N



Le chat mange

Encodeur

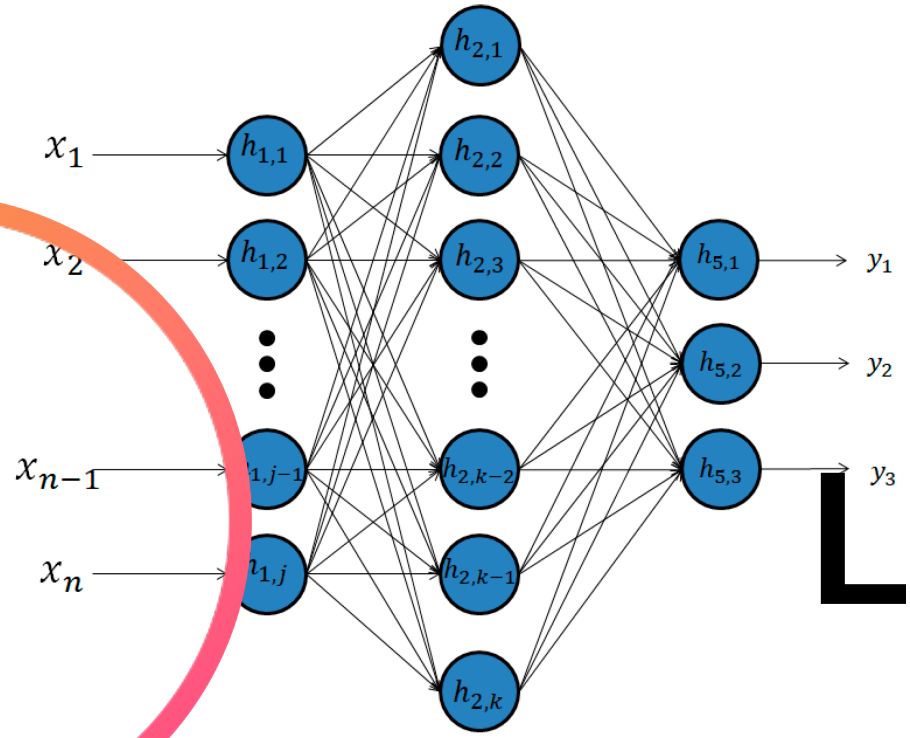
Décodeur



Recurrent Neural Network

Issue

Bohicon est une très belle ville du Benin, j’est vécu cinq et c’est là j’ai appris à parler …



Long short term memory