

# Data-Labs

“als vehicle voor het verantwoord inzetten van generatieve AI in het hoger onderwijs”

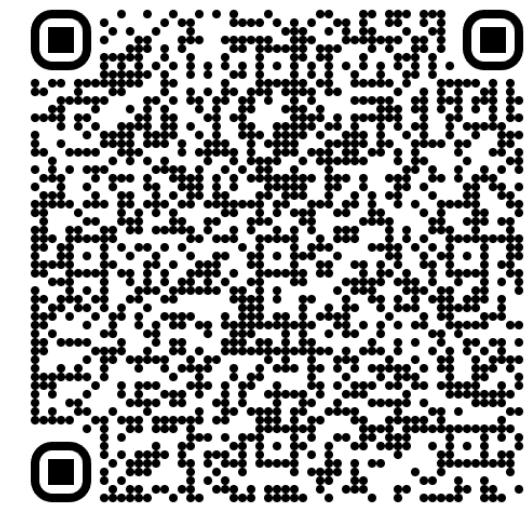
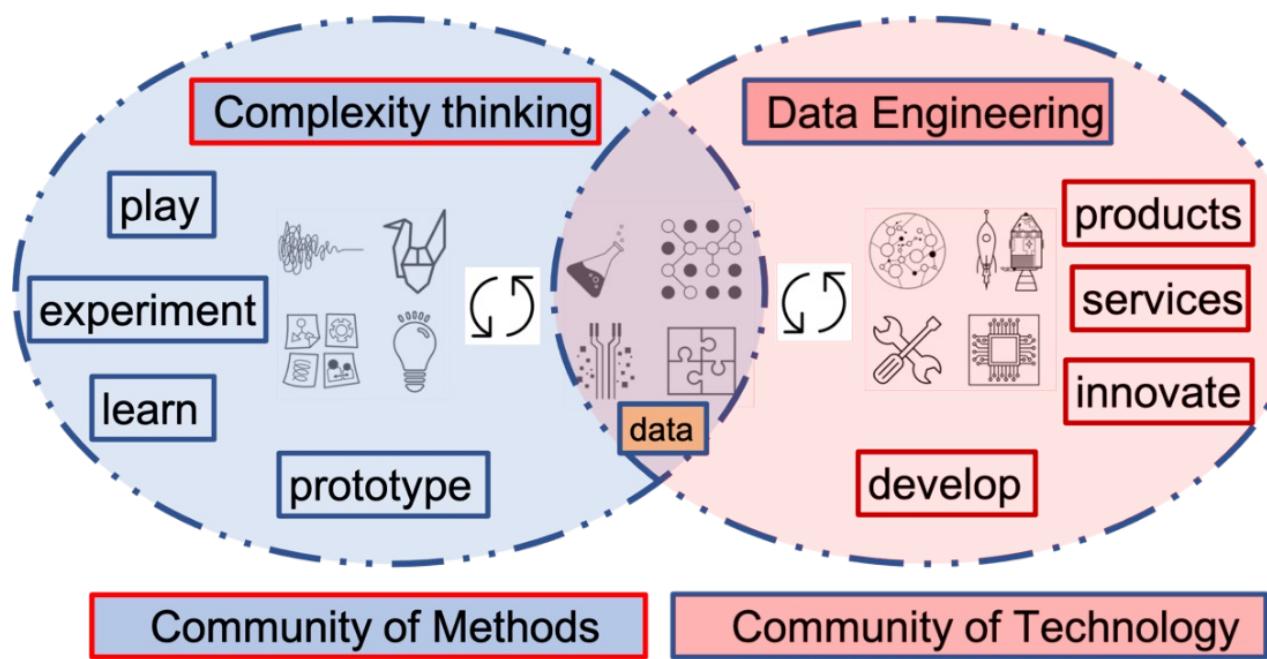
LunchLezing  
OKT 2023

Een open-source product gemaakt door het HR-brede programma voor AI & Ethisiek <http://hr.nl/ai>  
Rob van der Willigen

# CREATING A DATA FABRIC

Opbouwen van kennis & expertise  
door hands-on seminarie van datatechnologie

Artwork Adapted from: <https://idezo.ch>



<https://www.researchgate.net/publication/361164501> Creating a Data Fabric Through Easy-to-use Cloud Computing AiRA Luchlezing JUNI 2022

# Data-Lab Road-Map

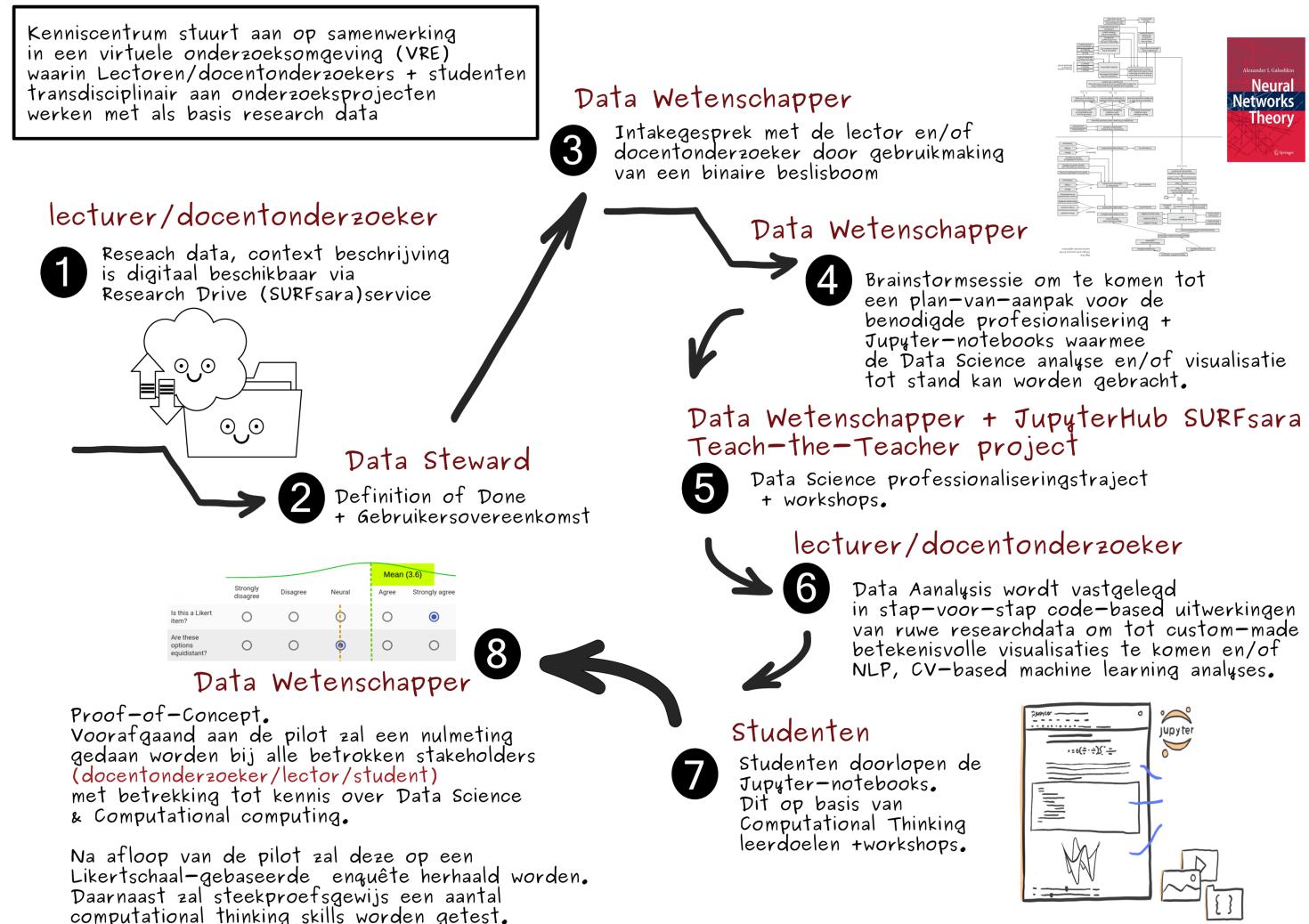
ResearchGate

Research Proposal PDF Available

Creating a Data Fabric through Easy-to-Use Cloud Computing. DCC SURF-Pilot 3de-ronde 2022. Produced by Living-Lab: AiRA, Hub voor Data & Responsible AI, Hogeschool Rotterdam. <https://www.surf.nl/en/news/six-new-pilots-awarded-in-dcc-for-practice-based-research>

May 2022  
DOI:10.13140/RG.2.2.25781.55526  
Affiliation: Hogeschool Rotterdam  
Project: AiRA [Artificial intelligence, Responsible and Applied]  
Authors:

 Robert Frans van der Willigen  
Hogeschool Rotterdam



## Discover Research Phase

- [0] Collection
- [1] Access + Retrieval

Big Data (Acquisition/Aggregation) Gathering  
Empirical(Sensor/IoT Measuring/Sampling)

Ownership (Open/Closed)  
Storage (Cloud/Database)



<https://robfvdw.medium.com/a-generic-approach-to-data-driven-activities-d85ad558b5fa>

## Data Gathering & Ingesting

## Define Synthesis Phase

- [2] Preparation + Wrangling (Munging)

Loading  
Feature Extraction/Reduction  
Normalization  
Transformation  
Conversion

## Develop Ideation Phase

- [3] Exploration
- [4] Analysis + Machine-Learning
- [5] Abstraction

Graphical (spatial)  
Ontological (Language)  
Semantic (text)  
Rule-based/Algorithmic  
Quantitative/Qualitative  
Numerical/Categorical/Symbolic

Mining (Heuristics/Statistics/Descriptive/Prescriptive)  
Construct Useful Insights/Trends/Patterns/Diagnosis(Information)

Parameter Selection + Representation  
Summarization  
Problem Solving  
Diagnostic  
Prediction  
Encription

## Deliver Implementation Phase

- [6] Organization + Managing
- [7] Automation + Reporting

Visualization  
Virtualization  
Performance (Measure/Monitor)  
Evaluation & Review  
Decision & Advise or Prescription  
(Interactive/Passive) Story Telling

## Data Dissemination & Curation

# Hoe bouw je Datalabs ?

## Proof-of-Concept

Toegevoegde waarde DataLabs voor de HR + KCs ...

“innovatie motor”

“digitale transitie”

“stimuleert kennisbenutting”

# **Waarom hebben we Data-labs nodig ?**

# Benutten van Onderzoeksdata vereist samenwerking

Onderzoekers besteden veel tijd aan het regelen van toegang tot de systemen en diensten die ze nodig hebben om hun onderzoek verantwoord te doen.

Onderzoekers kunnen vaak niet beschikken over een **directe toegang** tot de systemen en diensten (zoals virtuele samenwerking omgevingen) voor het gebruik van SotA Data Science Tools.

Onderzoekers kunnen vaak niet beantwoorden aan relevante regelgeving en standaarden om veilig te kunnen omgaan met researchdata.

# *Data infrastructure drivers*

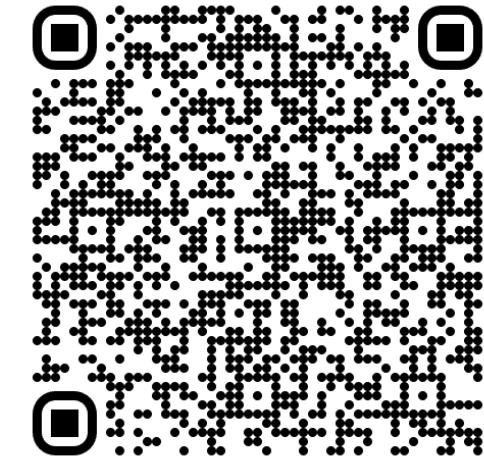
People

Data Access & Dissemination

Knowledge Semination & Data Curation

Interoperability, Cyber Security & Standards

Cloud Computing Capabilities & Accessibility



[https://www.researchgate.net/  
publication/357933768 NO  
MORE SECRETS AIRA Living-  
Lab AI ETHICS](https://www.researchgate.net/publication/357933768_NO_MORE_SECRETS_AIRA_Living-Lab_AI_ETHICS)

# Analyzing the Analyzers

An Introspective Survey of  
Data Scientists and Their Work  
Harlan D. Harris, Sean Patrick Murphy  
& Mack Vaisman



O'REILLY®

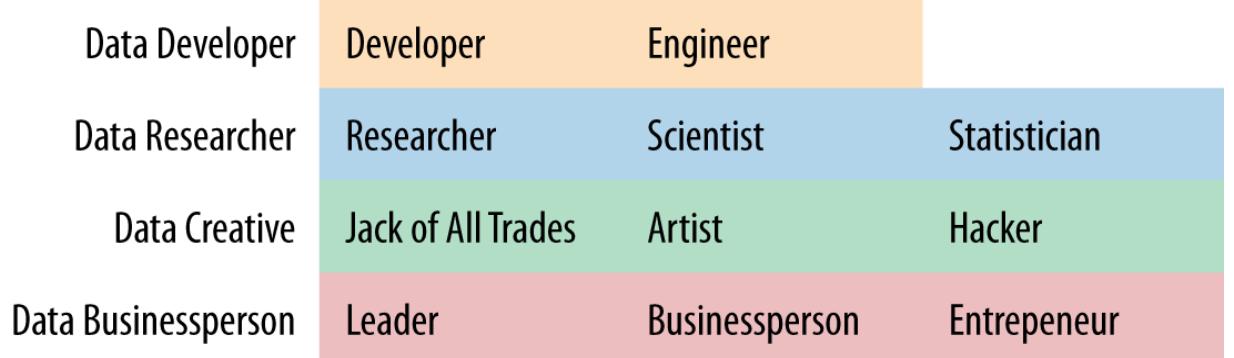
Strata

The size of your organization often determines role overlap



Business	ML / Big Data	Math / OR	Programming	Statistics
Product Development	Unstructured Data	Optimization	Systems Administration	Visualization
Business	Structured Data	Math	Back End Programming	Temporal Statistics
	Machine Learning	Graphical Models	Front End Programming	Surveys and Marketing
	Big and Distributed Data	Bayesian / Monte Carlo Statistics		Spatial Statistics
		Algorithms		Science
		Simulation		Data Manipulation
				Classical Statistics

[https://mlleo.github.io/ml/GCP\\_day1/](https://mlleo.github.io/ml/GCP_day1/)



# Domain Data and Analytics Model

Gartner.

## Hype Cycle for Compute Infrastructure, 2020

Published 8 July 2020 · ID G00448100 - 99 min read

By Analyst(s): Tony Harvey, Daniel Bowers, Chirag Dekate

Initiatives: [Data Center Infrastructure](#)

While AI, cloud and security remain hyped, COVID-19 instantly changed the priorities for I&O. For I&O leaders, this means urgently supporting the new imperatives around remote work and cost reduction while continuing to enable cloud, security and AI.

### Analysis

#### What You Need to Know

This document was revised on 2 October 2020. The document you are viewing is the corrected version. For more information, see the [Corrections](#) page on gartner.com.

Businesses were already challenging I&O to provide increased agility, rapid deployment of new applications and better performance. The COVID-19 pandemic has only increased the urgency, accelerating demand for compute innovations that can scale rapidly and deliver more, more cost-effectively. Gartner has identified the most relevant innovations in the compute space for I&O leaders to evaluate and deliver on these challenges.

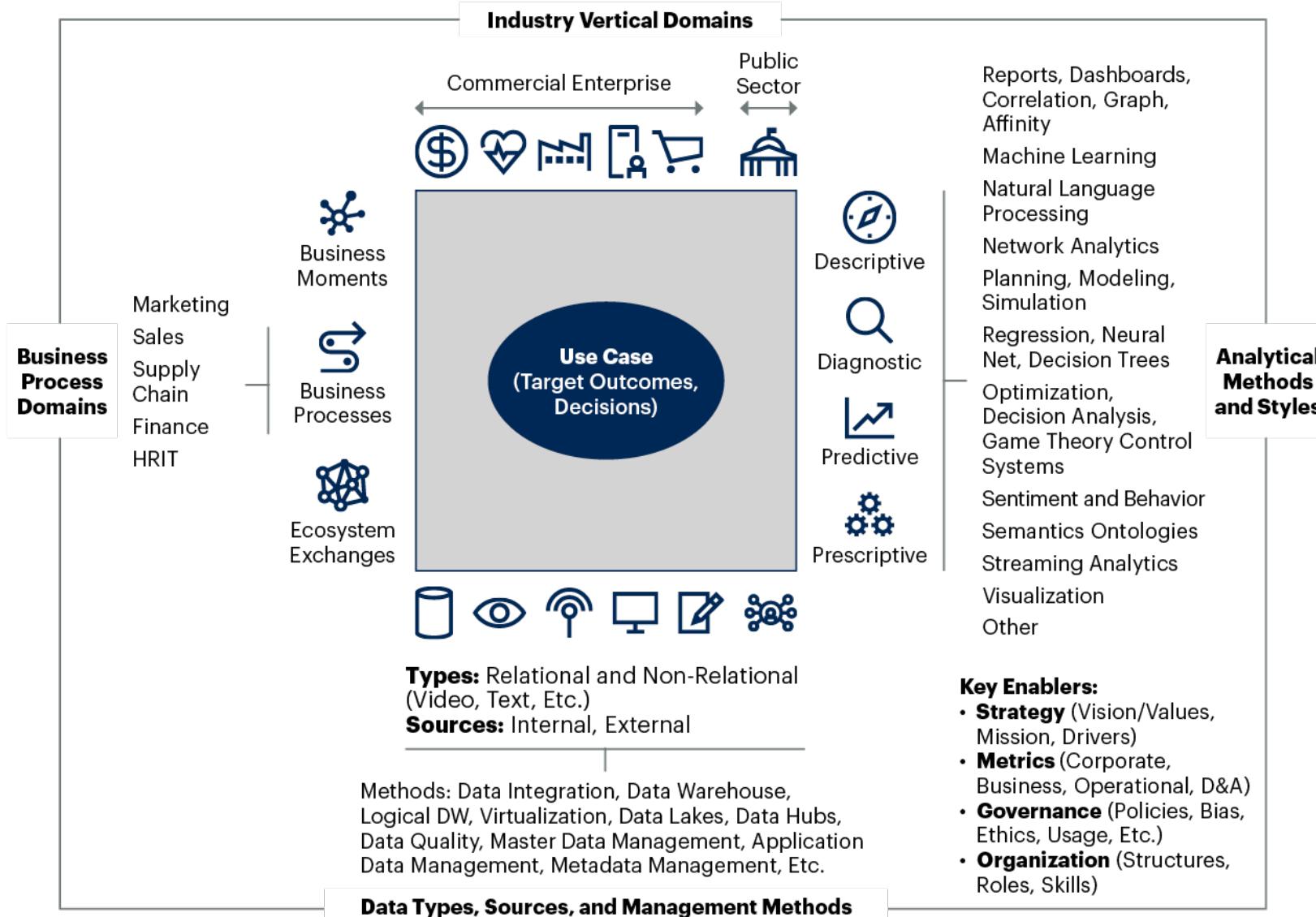
For more information about how peer I&O leaders view the technologies aligned with this Hype Cycle, please see "2020-2022 Emerging Technology Roadmap for Large Enterprises."

### The Hype Cycle

Compute infrastructure includes technologies used inside the data center and in off-premises locations, including cloud and edge.

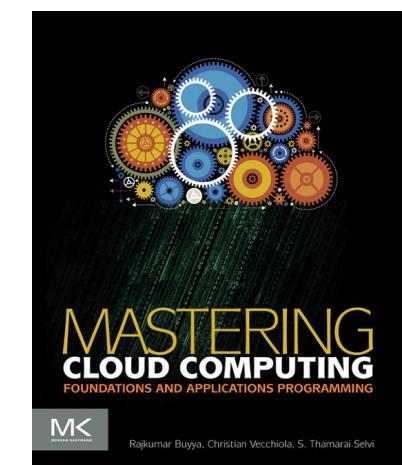
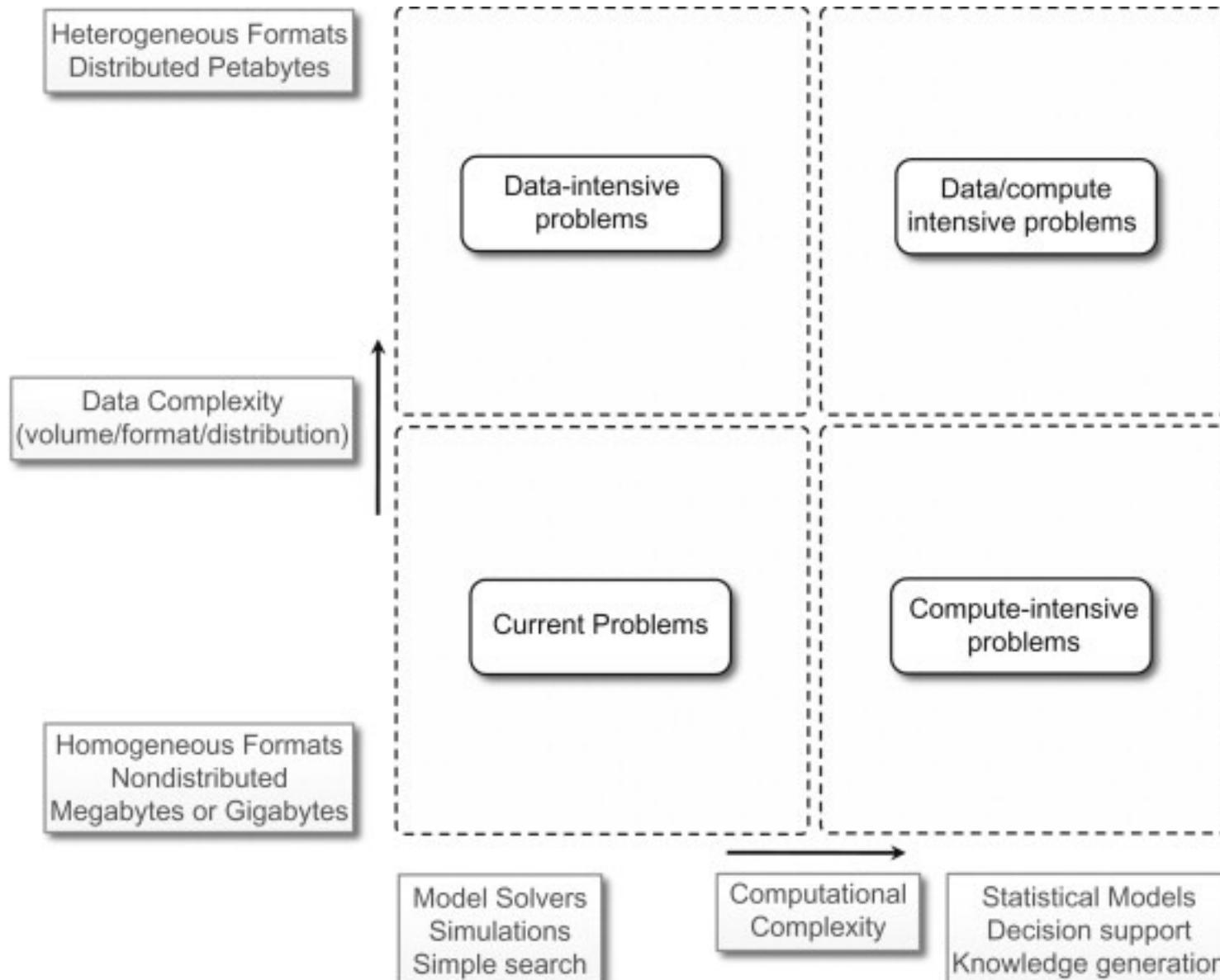
Cloud computing continues to transform I&O, and the rapid response required by the COVID-19 pandemic is driving more changes. Delivering IT services remotely will become the new normal for a significantly larger subset of end users. Strategic initiatives around migrating to cloud, remote work and automation will continue, and there will be a renewed focus on cost optimization.

Trends in compute infrastructure covered in this year's Hype Cycle include:

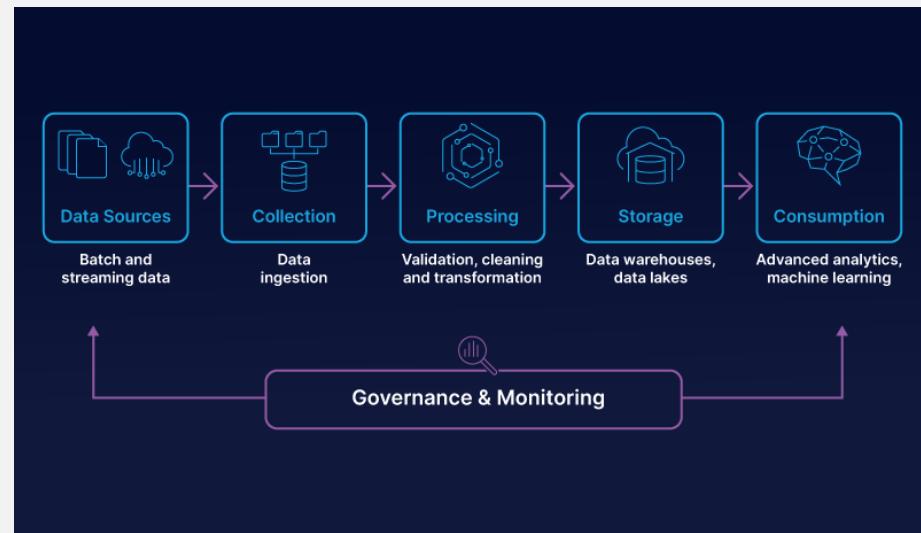


Source: Gartner

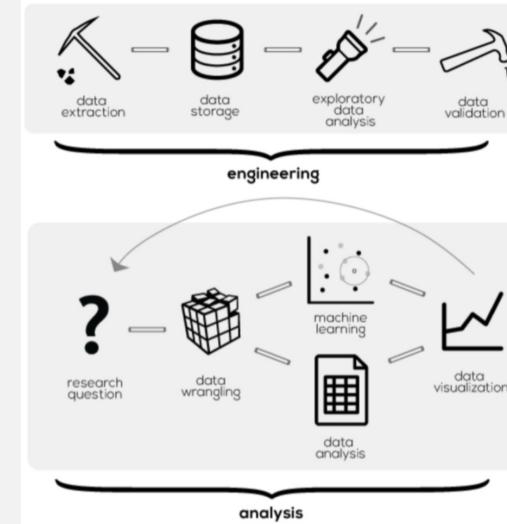
724483\_C



# Data pipelines zijn complex



## the data pipeline

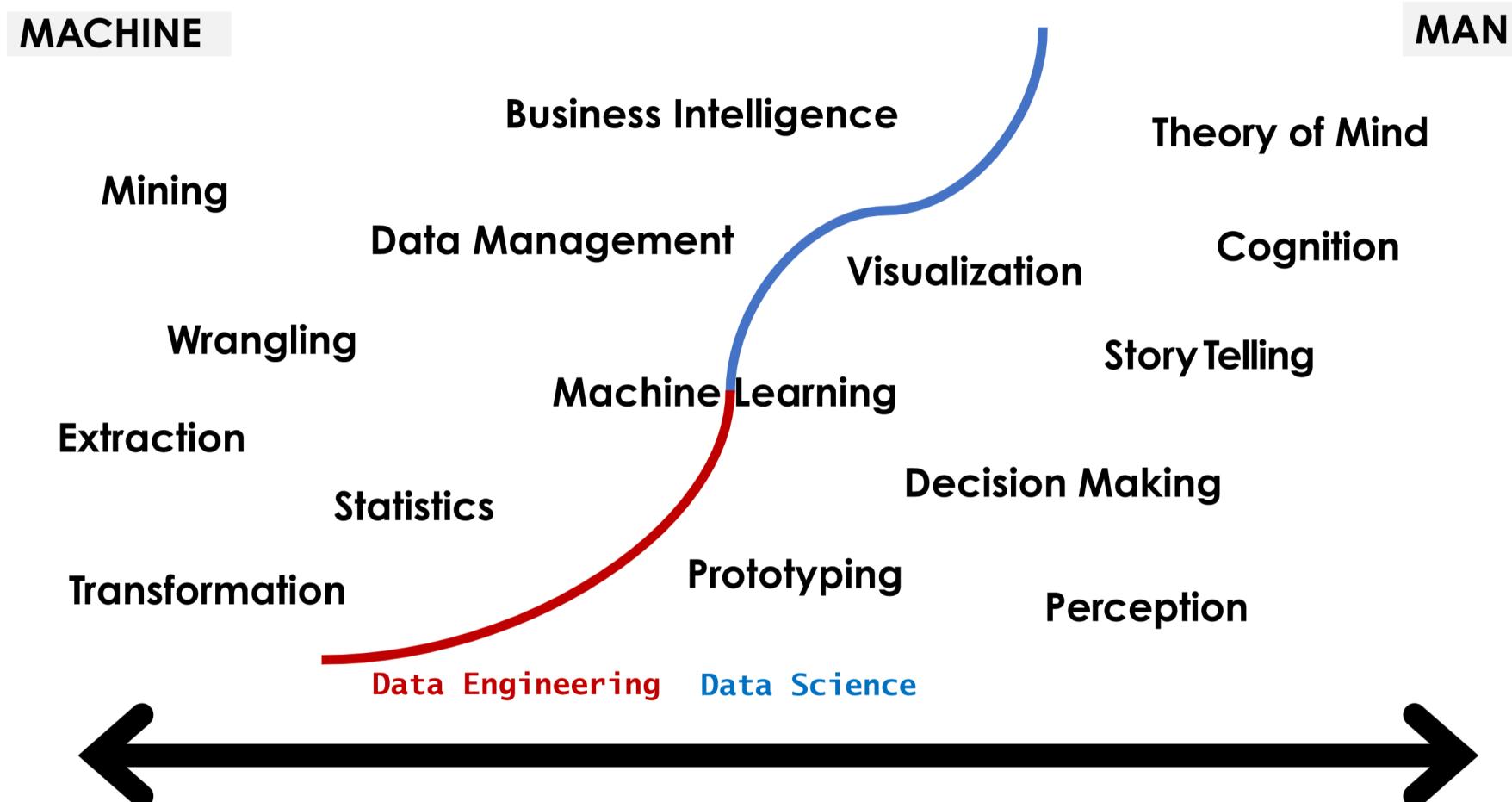


Data  
Engineering

Data  
Science

The concept of the data pipeline is depicted as outlined by Alex Petralia (2019)

# Werken met data vereist Engineering + Science



Inspired by Daniel Keim, "Visual Analytics: Definition, Process, and Challenges"

# **Stap (1)**

## **Toegang krijgen tot**

**Public, Local of Hybrid  
CLOUD COMPUTING**

# PUBLIC CLOUD COMPUTING

Public cloud computing is het beschikbaar stellen van schaalbare informatietechnologie resources (IT-resources) via het internet, in tegenstelling tot het lokaal hosten en beheren ervan.

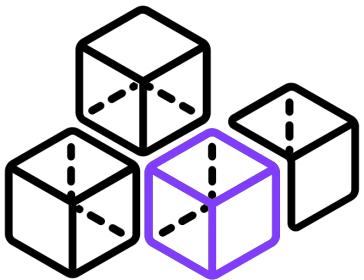
IT-resources kunnen softwareapplicaties en/of diensten omvatten, evenals de infrastructuur waarop ze werken.

Cloudcomputing maakt gebruik van zogeheten datacenters. Deze bestaan uit een conglomeraat aan hardware, software en netwerken die kunnen worden ingezet om clouddiensten te ontwikkelen, testen, leveren, bewaken, controleren of ondersteunen.

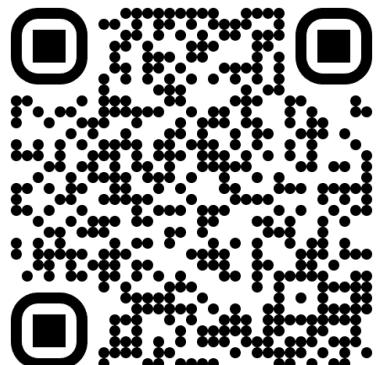
# Public Cloud Platforms



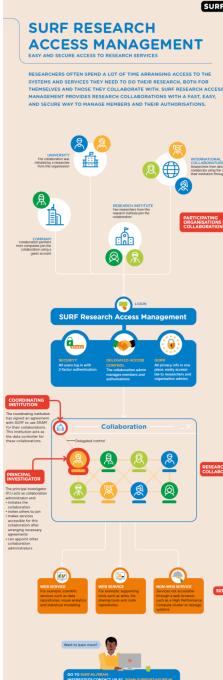
# Private Cloud: CMI-Datalab



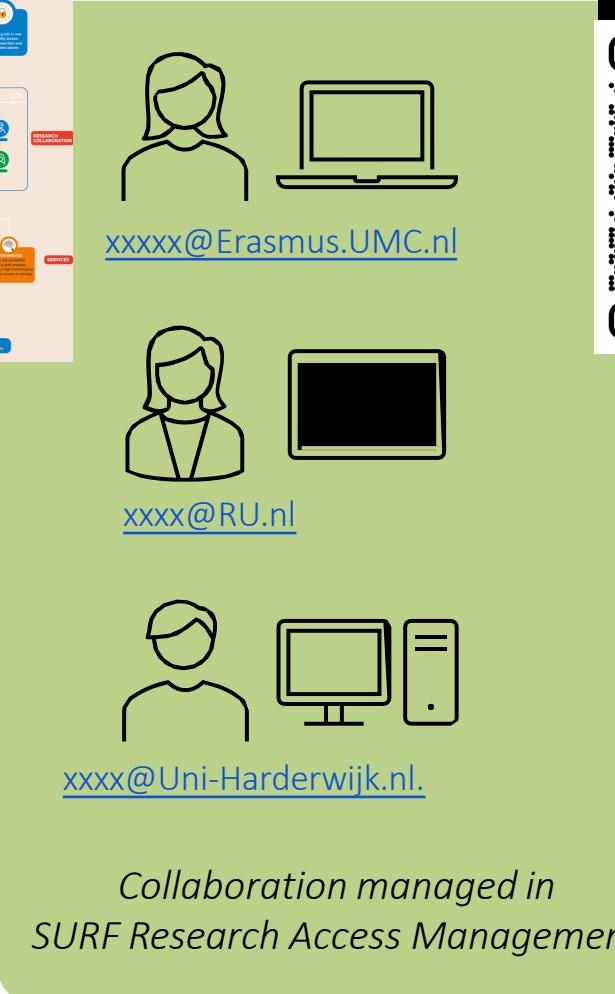
## CMI-Datalab Rotterdam



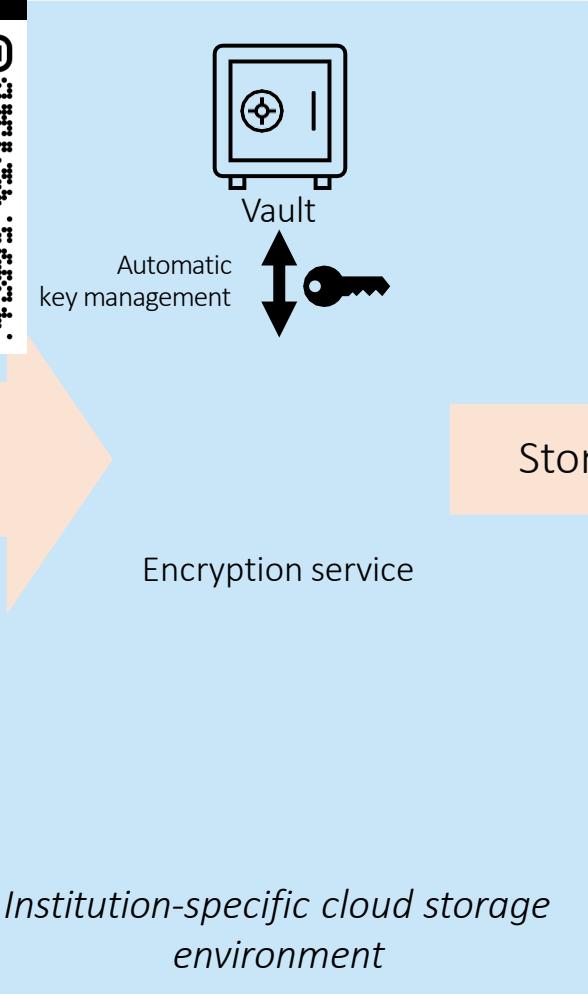
The screenshot shows a screenshot of the Datalab Rotterdam documentation website. The header includes the Datalab Rotterdam logo and links for "Services", "Docs", "Website", and "GitHub". The main content area is titled "Tutorials" and contains sections for "Introdutie", "Analytics", "Juyster", "MQTT", "Timeseries", "Tutorial", and "Internet of Things". Below this, there are sections for "Begrippenlijsten", "Veelgestelde Vragen", "Documentatie", "Hulpmiddelen en Bronnen", and "Ondersteuning nodig". Each section provides brief descriptions and links to further information.



# Hybride Cloud: **SRAM**



Login



## SECURITY

Inloggen met 2-factor authenticatie (FIDO2)

## GEDELEGEERDE TOEGANGSCONTROLE

Projectleider beheert login + machtigingen

Data Protection Impact Assessment (DPIA) + AVG-compliance

Alle privacy-informatie op één plaats  
voldoet aan alle standaarden en richtlijnen



# **Stap (2)**

## **Tool-Chain &**

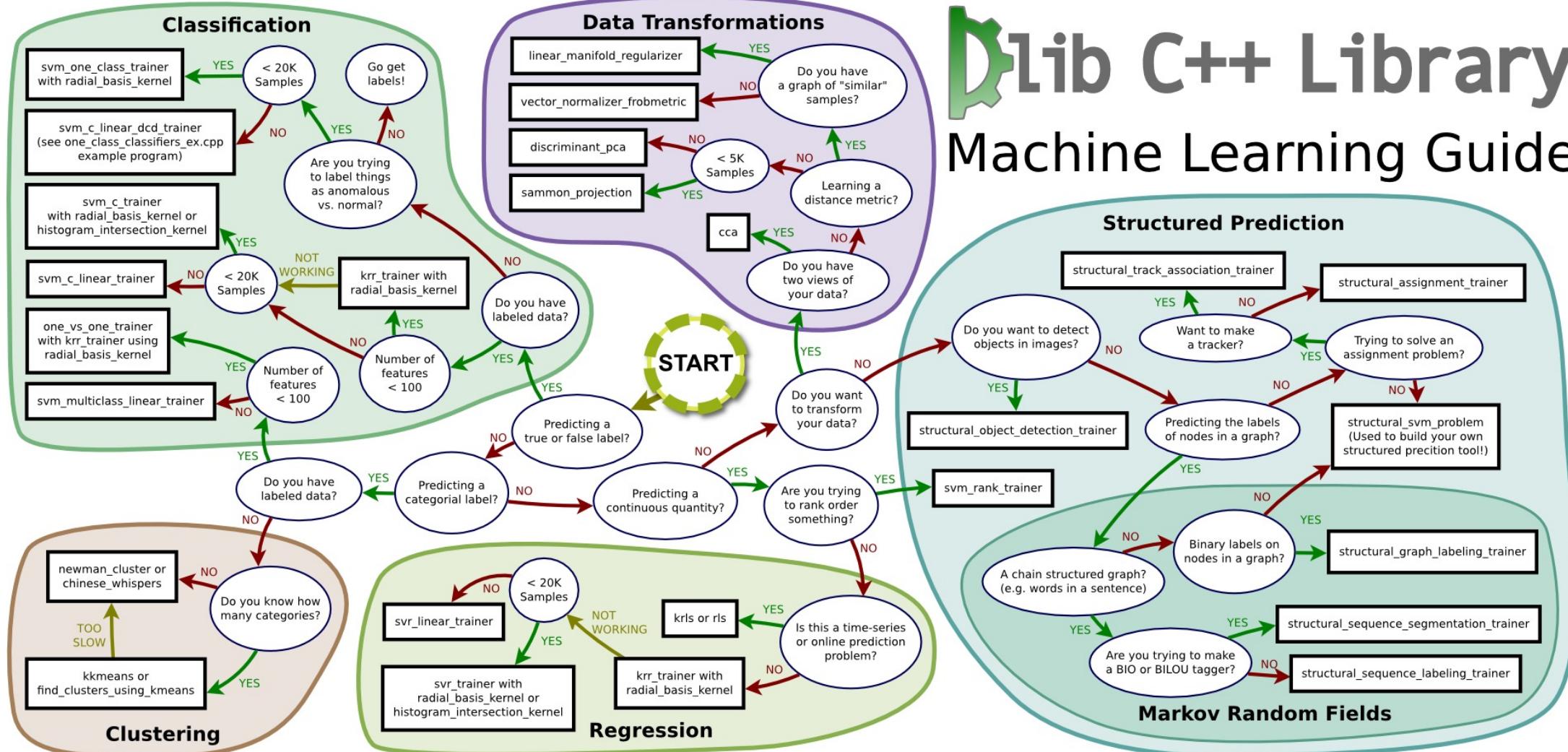
## **AI-algoritme selectie**

### **Software installatie**

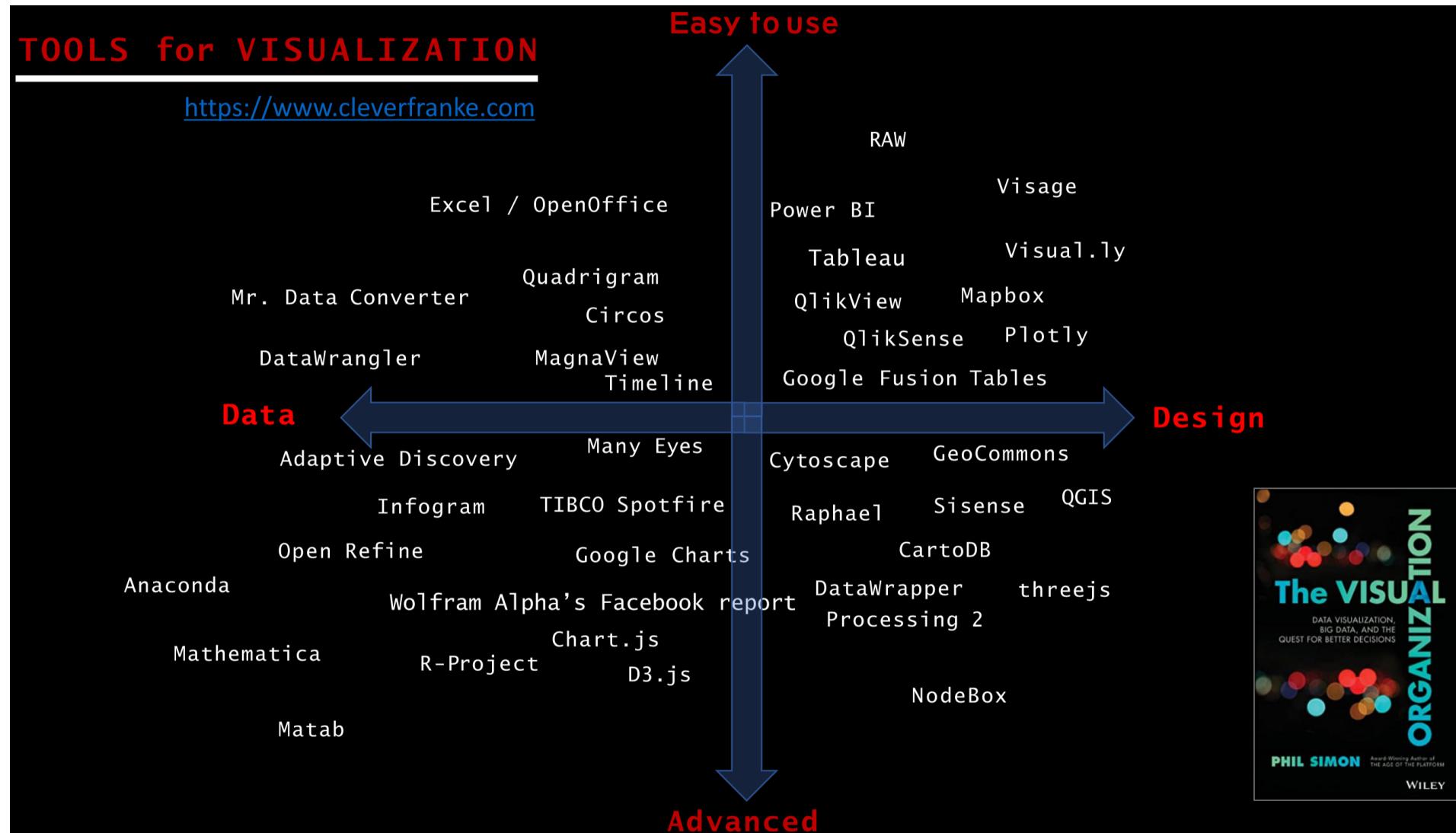
# {Software Installation is Complex}



HOGESCHOOL  
ROTTERDAM



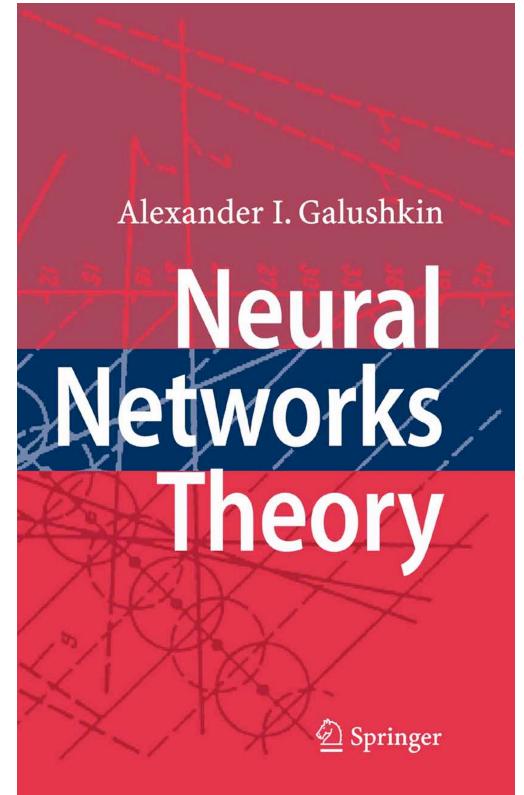
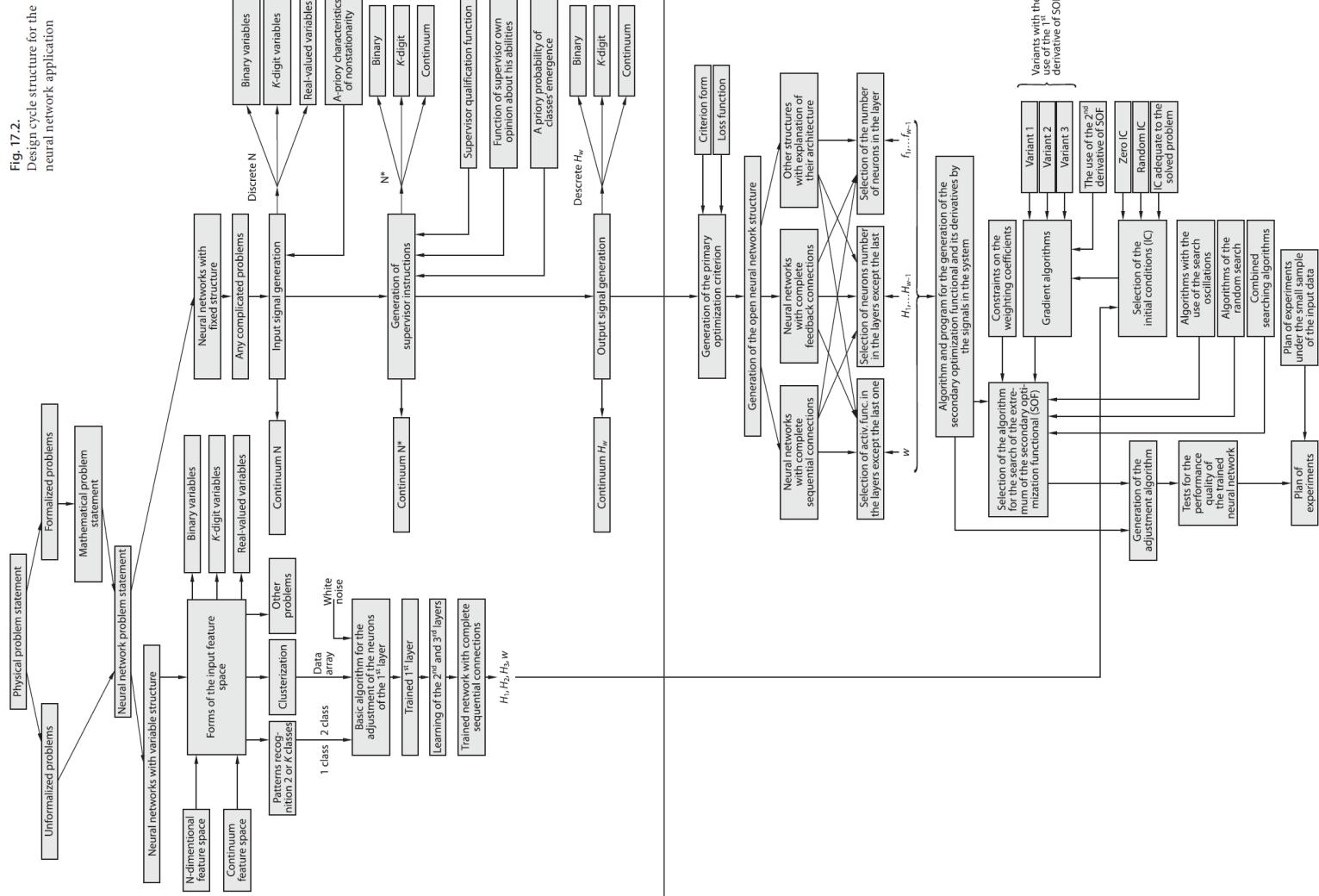
# {Cloud Computing → Easy Tooling Access}



HOGESCHOOL  
ROTTERDAM

# {Machine Learning }

Employ decision-three flowcharts to apply Machine Learning appropriately



# {Machine Learning}

## Requires Cloud Computing

• RESEARCH BRIEFS | DECEMBER 17, 2020

### Artificial Intelligence and the Future of Work

THOMAS MALONE, DANIELA RUS, ROBERT LAUBACHER

RB17-2020

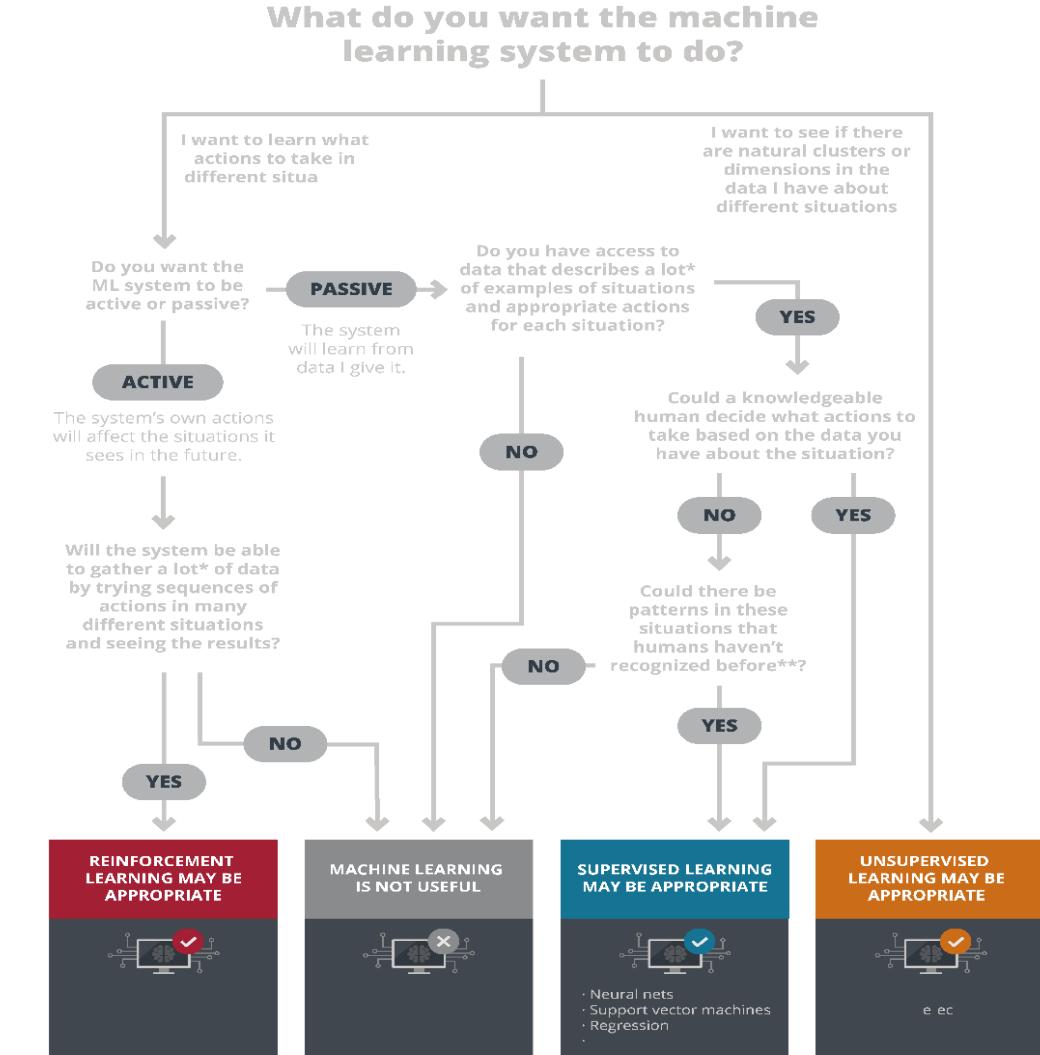


SHARE



DOWNLOAD RESEARCH BRIEF

<https://workofthefuture.mit.edu/wp-content/uploads/2020/12/2020-Research-Brief-Malone-Rus-Laubacher2.pdf>



# {Cloud Computing → access to AI frameworks}

Patterns

Tutorial  
Ready, Steady, Go AI: A practical tutorial  
on fundamentals of artificial intelligence  
and its applications in phenomics image analysis



Fabio Nistri and Antonio L. Marques<sup>a,b</sup>  
<sup>a</sup>Department for Innovation Biologics, Agro-food and Forest systems, University of Tuscia, Via S. Camillo de Lellis, Viterbo 01100, Italy  
<sup>b</sup>Corresponding author: fabio.nistri@unitus.it DOI: 10.31233/osf.io/2j11v33

THE BIGGER PICTURE Advances in AI technologies have the potential to significantly increase our ability to turn plant phenotypic data into valuable information. However, such analyses require specific programming skills and knowledge reserved for computer scientists. We provide an interactive tutorial with live, open-source, and FAIR notebooks that can allow researchers to conduct such analyses without the need for extensive programming experience. The tutorial is divided into two main parts: a first part that covers AI and XAI algorithms that augment and complement human experience in classifying tomato leaf diseases and spider mites. Our tutorial is not only applicable to other diseases but also transferable to other crops and species. It is also designed to be used in a classroom setting, so that students can learn by doing with their data. We expect our notebooks to be of high interest to those who want to enhance the performance and sustainability of agriculture.

Production Data science output is validated, understood, and regularly used for multiple domains/platforms

SUMMARY High-throughput image-based technologies are now widely used in the rapidly developing field of digital phenomics and are generating ever-increasing amounts of data. Artificial intelligence (AI) becomes a great tool for deriving the most value from this data in terms of disease detection and diagnosis. However, this requires specialized programming skills and an in-depth understanding of machine learning, deep learning, and XAI. This tutorial provides an introduction to the basic concepts of AI, the available development tools, technologies, and services available to the phenomics data community and show how they can be applied to selected problems in explainable AI-based image analysis. This tutorial provides practical and useful implementation examples and insights to harness the potential of the phenomics data in explainable AI and breeding programs.

HELLO, WORLD While this example is used as an illustration to teach about every programming language, it is also a good example to introduce C. In fact, here we incorporated instructions for older programming languages, such as the original version of C, which was developed by Dennis Ritchie in 1972. In 1978, he revised it in "The C Programming Language" book, which is still considered one of the best books on the subject. The original version of C is a low-level programming language, released in the style of low-level assembly language, and it is still used today.

Although PL/I was introduced in 1958, the model of AI date back to 1956, when Alan Turing, a British mathematician, gave a famous public lecture on computer intelligence. In 1956, this paper described the first AI program, called the "Turing Test". Five years later, Alan Newell, Cliff Shaw, and Herbert A. K. Hechtel wrote an AI program called the "Logic Theorist".

This is an open access article under the CC-BY-NC-ND license (<https://creativecommons.org/licenses/by-nd/4.0/>).

Table 1. Representative list of available open-source software frameworks and programming languages for the implementation of AI algorithms

Software framework	Programming languages	Operating systems	GPU computing support <sup>a</sup>	Distributed training support	Availability of pretrained models <sup>b</sup>	Software license	Website
<b>Deep ML frameworks<sup>c</sup></b>							
Tensorflow & Keras <sup>d</sup>	Python, JavaScript, C++, Java, Go, Swift (early release)	Linux, Windows, macOS	Linux, Windows	multiple GPUs, multiple machines	yes	Apache 2.0	<a href="http://www.tensorflow.org">www.tensorflow.org</a>
PyTorch	Python, C++, Java	Linux, Windows, macOS	Linux, Windows	multiple GPUs, multiple machines	yes	BSD-3	<a href="http://pytorch.org">pytorch.org</a>
Apache MXNet	Python, C++, Scala, Julia, Clojure, Java, R, Perl	Linux, Windows, macOS	Linux, Windows	multiple GPUs, multiple machines	yes	Apache 2.0	<a href="http://mxnet.apache.org">mxnet.apache.org</a>
H2O	Python, Scala, Java, R	Linux, Windows	Linux, Windows	multiple GPUs, multiple machines	no	Apache 2.0	<a href="http://www.h2o.ai">www.h2o.ai</a>
Deeplearning4j	Java, Groovy, Scala, Kotlin, Clojure	Linux, Windows, macOS, Android	Linux, Windows	multiple GPUs, multiple machines	yes	Apache 2.0	<a href="http://deeplearning4j.org">deeplearning4j.org</a>
Chainer	Python	Linux	Linux	multiple GPUs, multiple machines	yes	MIT	<a href="http://chainer.org">chainer.org</a>
PaddlePaddle	Python	Linux, Windows, macOS	Linux, Windows	multiple GPUs, multiple machines	yes	Apache 2.0	<a href="http://www.paddlepaddle.org">www.paddlepaddle.org</a>
SINGA	Python, C++	Linux, Windows, macOS	Linux, Windows	multiple GPUs, multiple machines	no <sup>e</sup>	Apache 2.0	<a href="http://singa.apache.org">singa.apache.org</a>
Flux	Julia	Linux	Linux	multiple GPUs	no	MIT	<a href="http://fluxml.ai">fluxml.ai</a>
OpenNN	C++	Linux, Windows, macOS	Linux, Windows	no	no	LGPLv3	<a href="http://www.opennn.net">www.opennn.net</a>
Dlib	Python, C++	Linux, Windows, macOS	Linux, Windows	multiple GPUs	no	BSL-1.0	<a href="http://dlib.net">dlib.net</a>
MLBox	Python	Linux, Windows, macOS	no	no	no	BSD-3	<a href="http://github.com/AxeldeRombay/MLBox">github.com/AxeldeRombay/MLBox</a>



HOGESCHOOL  
ROTTERDAM

**Stap (3)  
benutten van**

**Jupyter Hub Notebooks  
Research Drive**

# {JupyterHub → on-demand notebooks}

JOURNAL OF STATISTICS AND DATA SCIENCE EDUCATION  
2021, VOL. 29, NO. S1, S103–S111  
<https://doi.org/10.1080/10691898.2020.1860726>

Taylor & Francis  
Taylor & Francis Group

OPEN ACCESS

Check for updates

## Easy-to-Use Cloud Computing for Teaching Data Science

Brian Kim<sup>a</sup> and Graham Henke<sup>b</sup>

<sup>a</sup>Joint Program in Survey Methodology, University of Maryland, College Park, MD; <sup>b</sup>Center for Urban Science & Progress, New York University, Brooklyn, NY

**ABSTRACT**  
One of the biggest hurdles of teaching data science and programming techniques to beginners is simply getting started with the technology. While multiple versions of the same coding language exist (e.g., Python 2 vs Python 3), numerous additional libraries and packages to install, as well as integrated development environments to navigate, the first step can be the most daunting. We show the advantages of using cloud computing to solve this issue and demonstrate one way of implementing it to allow beginners to get started with coding immediately. Using user-friendly Jupyter notebooks along with the interactive capabilities possible through Binder, we provide introductory Python and SQL material that students can access without downloading anything. This lets students to get started with coding right away without getting frustrated figuring out what to install. Example introductory modules on using Python and SQL for data analysis are provided through GitHub at <https://github.com/Coleridge-Initiative/ada-intro-python> and <https://github.com/Coleridge-Initiative/ada-intro-sql>.

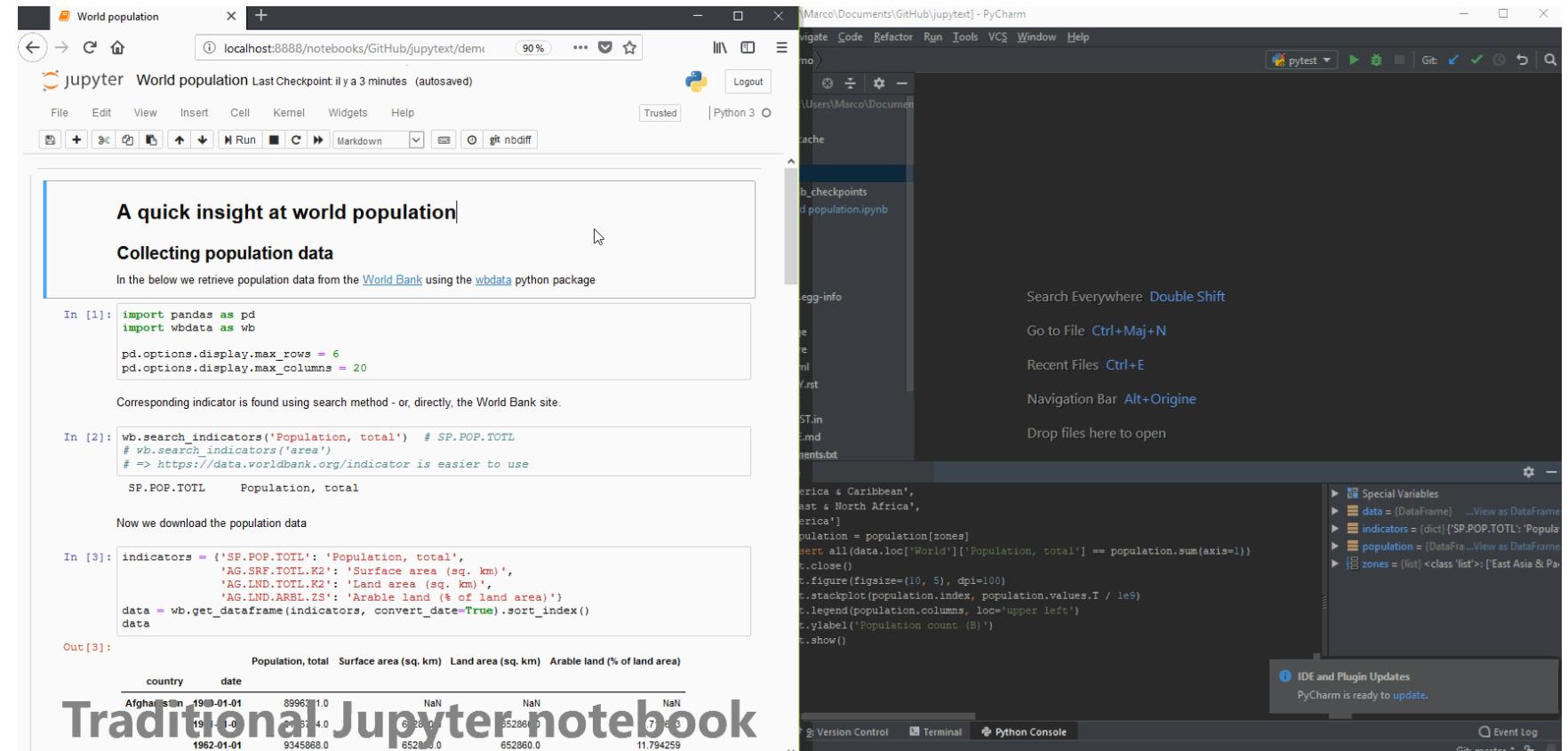
**KEYWORDS**  
Education; Jupyter; Python; SQL; Training

### 1. Introduction

The big data revolution has resulted in a great need for new ways to teach data science tools. A large part of the improvements in data science education has been in revamping current curricula, introducing new data science courses as well as establishing major and degrees. Additionally, many tools have been developed and refined to facilitate traditional classroom learning of the modern data analytics workflow, such as incorporating version control with GitHub Classroom (Fiksel et al. 2019). In particular, the social sciences have lagged behind in keeping up with more modern statistical methods (Aiken et al. 1990; King 2011). Many efforts have been made to improve undergraduate and graduate education by targeting how the curricula are built to support data science (Nolan and Temple Lang 2010; Baumer 2015; Ebyers and Hattingh 2016; Aikat et al. 2017; Cetinkaya-Rundel and Rundel 2018; Loy, Kuiper, and Chihara 2019; Yan and Davis 2019), and these efforts have extended to even secondary education (Gould et al. 2016).

However, the effort toward modernizing data analytics goes beyond the traditional classroom setting. Much of the focus has been on improving data science coursework and curricula, but an oft-overlooked segment of the population is working professionals such as staff at federal statistical agencies (Farmin et al. 2014). Data science education is just as relevant for professionals who are already part of the workforce and have been working on important problems using restricted data already (Kreuter, Ghani, and Lane 2019). The challenges of building up professional development have been extensively discussed in a National Academies Roundtable on Data

In this article, we aim to demonstrate the ease with which cloud computing can be implemented to teach programming. First, in Section 2, we discuss the motivation behind using a cloud-based environment along with the benefits and drawbacks within the context of the type of course or program



A screenshot of a Jupyter notebook interface titled "World population". The notebook shows a section titled "A quick insight at world population" and "Collecting population data". It includes code snippets for importing pandas and wbdata, searching indicators, and displaying population data. Below the code, a table is shown with columns: country, date, Population, total, Surface area (sq. km), Land area (sq. km), and Arable land (% of land area). The table data is as follows:

country	date	Population, total	Surface area (sq. km)	Land area (sq. km)	Arable land (% of land area)
Afghanistan	1990-01-01	899671.0	Nan	Nan	Nan
	1991-1-0	9114.0	6.287	652866.0	7.63
	1992-01-01	9345868.0	652866.0	652860.0	11.794259

The PyCharm IDE window on the right shows a file named "population.ipynb" with code related to population data analysis.

# {JupyterHub → many available Kernels}

<https://github.com/jupyter/jupyter/wiki/Jupyter-kernels>

[jupyter / jupyter](#) Public

<> Code Issues 185 Pull requests 6 Actions Projects Wiki Security Insights

## Jupyter kernels

Seçkin Kükrer edited this page 8 days ago · 154 revisions

### Jupyter kernels

Kernel Zero is IPython, which you can get through ipykernel, and is still a dependency of jupyter. The IPython kernel can be thought of as a reference implementation, as CPython is for Python.

Here is a list of available Jupyter kernels. If you are writing your own kernel, feel free to add it to the table!

Name	Jupyter/IPython Version	Language(s) Version	3rd party dependencies	Example Notebooks
Micronaut		Python>=3.7.5, Groovy>3	Micronaut	<a href="https://github.com/stainlessai/microne_jupyter/blob/master/examples/basic-service/notebooks/use-library.ipynb">https://github.com/stainlessai/microne_jupyter/blob/master/examples/basic-service/notebooks/use-library.ipynb</a>
Agda kernel		2.6.0		<a href="https://mybinder.org/v2/gh/clem/agd/kernel/master?filepath=example/Lab1mp.ipynb">https://mybinder.org/v2/gh/clem/agd/kernel/master?filepath=example/Lab1mp.ipynb</a>
Dyalog Jupyter Kernel		APL (Dyalog)	Dyalog >= 15.0	Notebooks
Coarray-Fortran	Jupyter 4.0	Fortran 2008/2015	GFortran >= 7.1, OpenCoarrays, MPICH >= 3.2	Demo, Binder demo
LFortran				Binder demo
Ansible Jupyter Kernel	Jupyter 5.6.0.dev0	Ansible 2.x		Hello World
sparkmagic	Jupyter >=4.0	Pyspark (Python 2 & 3), Spark (Scala), SparkR (R)	Livy	Notebooks, Docker Images

**Interface Level**

**Kernel Level**

Kernel	Jupyter Version	Python Version	Dependencies
sas_kernel	Jupyter 4.0	python >= 3.3	SAS 9.4 or higher
IPyKernel	Jupyter 4.0	python 2.7, >= 3.3	pymq
Julia		julia >= 0.3	
iHaskell		ghc >= 7.6	

NOTE: This is just an approximation

tslab		typescript 3.7.2, JavaScript ESNext	Node.js	Example notebooks				
IJavascript		nodejs >= 0.10						
ITypeScript		TypeScript >= 2.0	Node.js >= 0.10.0					
jpCoffeescript		coffeescript >= 1.7						
jp_LiveScript		livescript >= 1.5						
ICSharp	Jupyter 4.0	C# 4.0+	scriptcs					
IRKernel	IPython 3.0	R 3.2	rzmq					
SageMath	Jupyter 4	Any	many					
pari_jupyter	Jupyter 4	PARI/GP >= 2.9						
IFSharp	Jupyter 4	F#		Features				
Igo	Jupyter >= 4, JupyterLab	Go >= 1.8	ZeroMQ (4.x)	Example				
iGalileo	Jupyter >= 4, JupyterLab	Galileo >= 0.1.3						
gopherlab	Jupyter 4.1, JupyterLab	Go >= 1.6	ZeroMQ (4.x)	examples				
Gophernotes	Jupyter 4, JupyterLab, interact	Go >= 1.9	ZeroMQ 4.xx	examples				
Go		Go >= 1.4						
IScala		Scala						
almond (old name: Jupyter-scala)	IPython>=3.0	Scala>=2.10		examples				
IErlang	IPython 2.3	Erlang	rebar					
ITorch	IPython >= 2.2 and <= 5.x	Torch 7 (LuaJIT)						
Elixir	Jupyter >= 4.0	Elixir >= 1.5	Erlang OTP >= 19.3, Rebar	example, Boyle package manager examples, Boyle examples with usage Matrix library				
ierl	Jupyter >= 4.0	Erlang >= 19, Elixir >= 1.4, LFE 1.2	Erlang, (optional) Elixir					
IAldor	IPython >= 1	Aldor						
IOCaml	IPython >= 1.1	OCaml >= 4.01	opam					
IWolfram							Wolfram Mathematica(R), Metakernel	
Lua Kernel							Lua	
IPurescript							Purescript	
IPyLua							Lua	
ILua							Lua	

<https://github.com/TwistedHardware/mltutorial/blob/master/notebooks/jupyter/1.%20Introduction.ipynb>

# {JupyterHub → online service desks}

SURF Spaces

https://servicedesk.surf.nl/wiki/pages/viewpage.action?pageId=30660252

Search Log in

Research Drive

Snellius and Lisa

- FAQ Snellius/Lisa
- Cartesius to Snellius migration
- Obtaining an account
- Snellius**
- Lisa
- HPC User Guide
- Tutorials (HPC)
- Access Control Lists (ACLs)
- Ansys on Snellius
- Code Optimisation
- Deep Learning on A100 GP
- EasyBuild tutorial
- Environment Modules
- First time usage of Lisa
- gperf tools tutorial
- Jupyter Notebooks on Snellius**
- MPI tutorial
- 'myquota' end-user tooling
- Python tutorial
- R tutorial
- Running a remote visualization
- Singularity (to become App)
- Two-factor authentication
- Using project space for shared files
- Using Visual Studio Code for C/C++
- Using PyCharm for remote development
- Using STOPOS
- Valgrind tutorial
- OpenCL example
- Scheduling and backfilling
- Creating an SSH key pair without a password
- Software
- Spider

Pages / ... / Tutorials (HPC)

## Jupyter Notebooks on Snellius/Lisa

**Synopsis**

This documentation is for researchers on Snellius or Lisa who would like to use Jupyter Notebooks for their research. If you are a teacher using our Jupyter for Education service on Lisa for a course, please refer to the documentation [for that service](#).

- Running Jupyter Notebooks using Lisa's JupyterHub
  - Starting a Jupyter Notebook Server using Lisa's JupyterHub
  - The software environment on Lisa's JupyterHub
  - Customizing the module environment on Lisa's JupyterHub
  - Using Python virtual environments with Lisa's JupyterHub
  - Using Conda virtual environments with Lisa's JupyterHub
- Running Jupyter Notebooks on Lisa/Snellius using your own batch script
  - Step 1: batch job for launching a Jupyter Notebook Server
  - Step 2: create a reverse tunnel to the login node
  - Step 3: connect to the running Jupyter Notebook Server

### Running Jupyter Notebooks using Lisa's JupyterHub

You can start your own Jupyter Notebook through the instructions in [Running Jupyter Notebooks on a batch system using your own batch script](#). However, on Lisa, we run a JupyterHub, which can start a Jupyter Notebook Server for you and will connect you to it. This is generally much easier.

### Starting a Jupyter Notebook Server using Lisa's JupyterHub

To start a Jupyter Notebook on Lisa using the JupyterHub:

- Go to the web address of the JupyterHub (e.g. <https://jupyter.lisa.surfsara.nl/2021> if you want to run based on the '2021' module environment. The JupyterHub running with the previous environment, <https://jupyter.lisa.surfsara.nl/2020> will also remain available for the foreseeable future).
- Login using your regular Lisa credentials.
- Select from the dropdown menu if you want to start a Jupyter Notebook on a CPU or GPU node and click 'Start'. Note that Jupyter Notebook jobs started through the hub will be accounted like any other normal job, and that GPU nodes are more expensive.

JupyterHub will now submit a batch job for you in the background (you can see it using the 'squeue' command). Once the Jupyter Notebook Server comes online, the JupyterHub page will forward automatically and you'll be in the Jupyter Notebook environment.

Note that we keep some limited, but dedicated resources free to run these notebooks so that they start quickly. In general, this means that the Jupyter Notebook Server will either start 'immediately' (with 1 minute) or the job will time out (if the dedicated resources are filled up already by other users). If the job times out, please try again later when fewer users are using the system.

### The software environment on Lisa's JupyterHub

The software environment in which Lisa's JupyterHub starts contains a predetermined list of modules that are loaded. As an example: <https://jupyter.lisa.surfsara.nl/2020> will load R and Python modules from the '2020' module environment on Lisa, while <https://jupyter.lisa.surfsara.nl/2021> will load R and Python modules from the '2021' module environment. To see exactly which modules are loaded:

- Start a Jupyter Notebook session
- Check 'Terminal'

<https://servicedesk.surf.nl/wiki/pages/viewpage.action?pageId=30660252>

# {JupyterHub & Reproducibility}

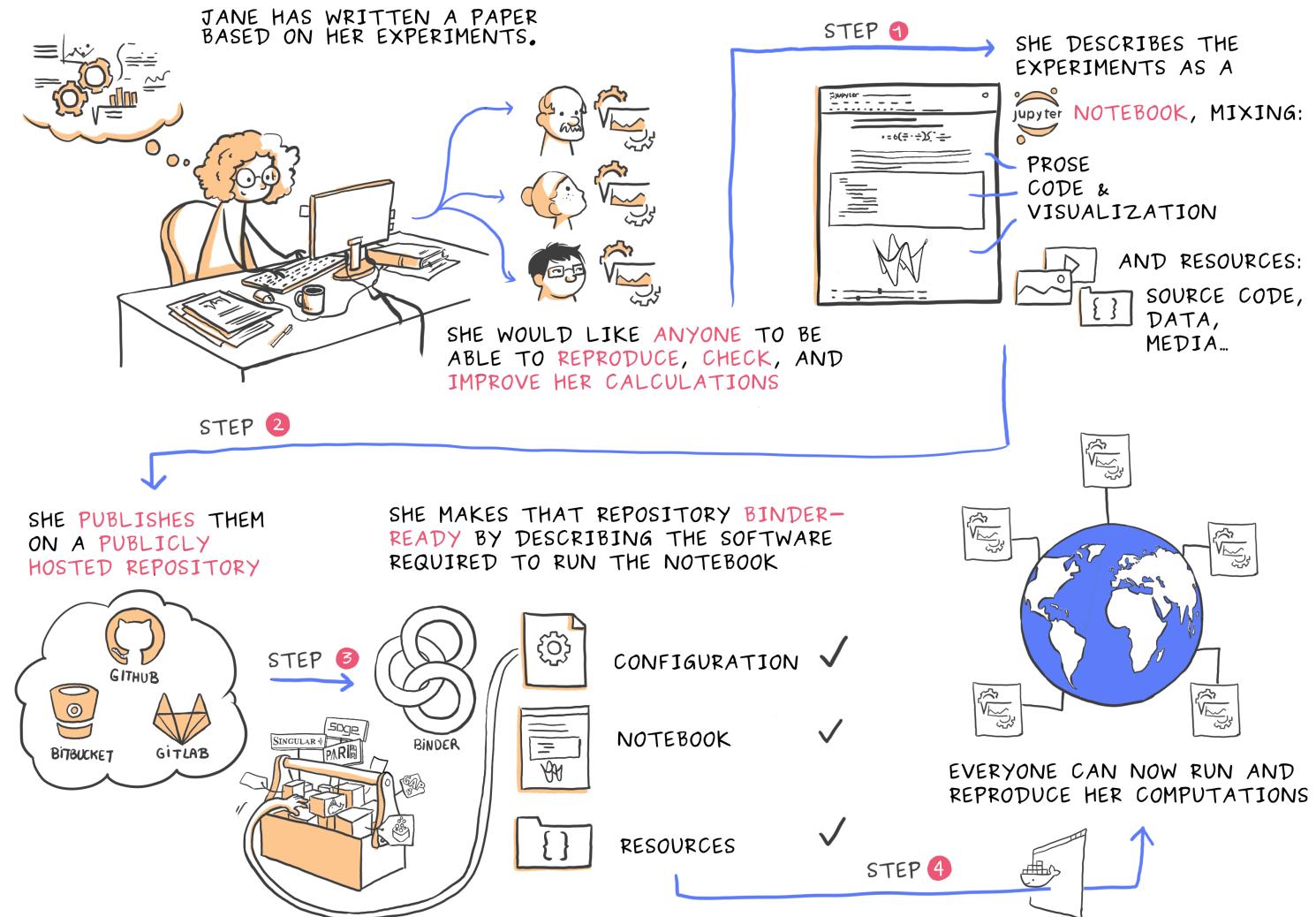


Literate computing has emerged as an important tool for computational studies and open science, with growing folklore of best practices. In this work, we report two case studies—one in computational magnetism and another in computational mathematics—where domain-specific software was exposed to the Jupyter environment. This enables high level control of simulations and computation, interactive exploration of computational results, batch processing on HPC resources, and reproducible workflow documentation in Jupyter notebooks. In the first study, Ubermag drives existing computational micromagnetics software through a domain-specific language embedded in Jupyter. In the second, a custom Jupyter kernel interfaces with the GAP system for computational discrete algebra and a declarative programming language. In light of these case studies, we discuss the benefits of this approach, including progress toward more reproducible and reusable research results and outputs, notably through the use of infrastructure such as Jupyterhub and Binder.

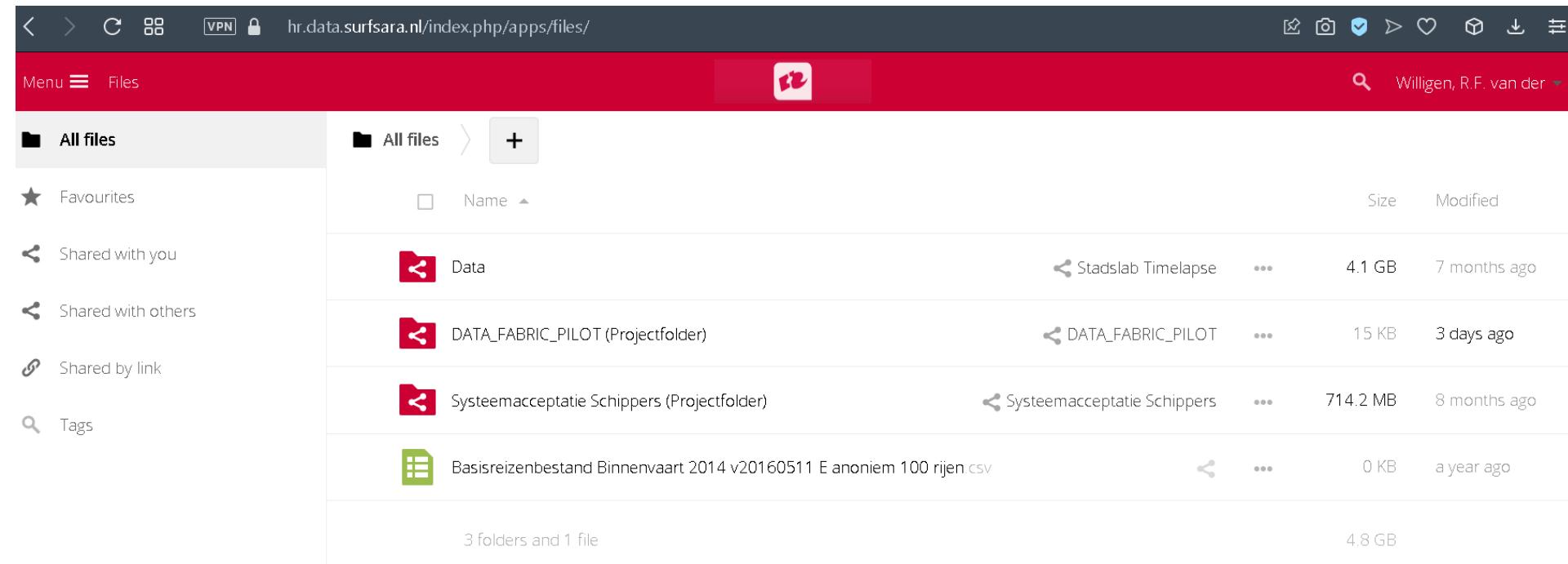
Research usually results in a publication that presents and shares the obtained findings and conclusions. For a publication to be scientifically valid, it must present the methodology rigorously, so that readers can follow the "recipe" and reproduce the results. If these details in the publication are considered reproducible, reproduced publications are more easily reusable and, thus, provide a significant opportunity to make (often tax-payer funded) research more transparent. However, the reproducibility of computational work is often hindered not only by a lack of data or metadata but also by a lack of details on the procedure and tools used.

10.1109/ICSE5109.2021.952101  
Digital Object Identifier: 10.1109/ICSE5109.2021.952101  
Date of publication 15 January 2021; date of current version 25 March 2021.

36 Computing in Science & Engineering Published by the IEEE Computer Society March/April 2021



# {HR SURFSARA Research Drive}



The screenshot shows a web-based file management interface. The top navigation bar includes icons for back, forward, search, and user information (Willigen, R.F. van der). The address bar shows the URL: <https://hr.data.surfsara.nl/index.php/apps/files/>. The left sidebar has a 'Menu' section with 'Files' selected, and links for 'All files', 'Favourites', 'Shared with you', 'Shared with others', 'Shared by link', and 'Tags'. The main content area displays a list of files and folders:

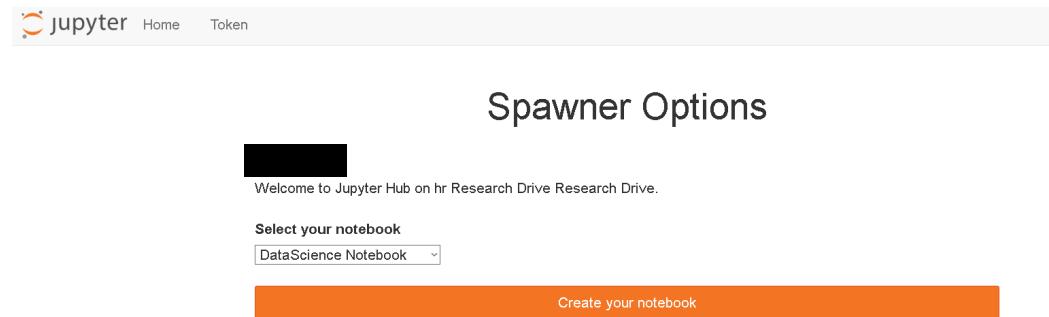
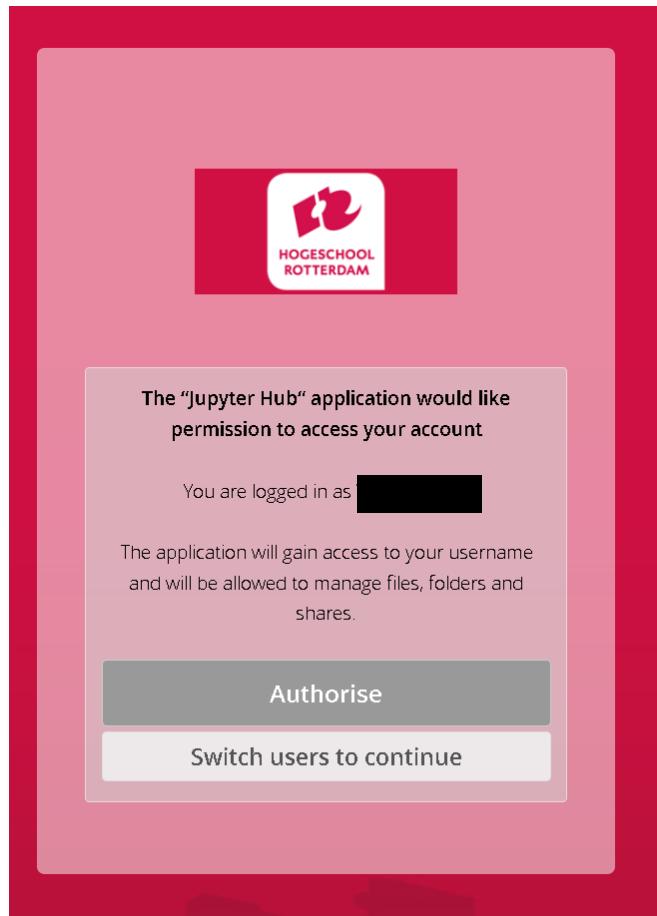
	Name	Owner	Size	Modified
	Data	Stadslab Timelapse	4.1 GB	7 months ago
	DATA_FABRIC_PILOT (Projectfolder)	DATA_FABRIC_PILOT	15 KB	3 days ago
	Systeemacceptatie Schippers (Projectfolder)	Systeemacceptatie Schippers	714.2 MB	8 months ago
	Basisreizenbestand Binnenvaart 2014 v20160511 E anoniem 100 rijen.csv		0 KB	a year ago

Total summary at the bottom: 3 folders and 1 file, 4.8 GB.

<https://hr.data.surfsara.nl/index.php/login>

# {HR SURFSARA Research Drive}

Sign in with your hr Research Drive account



The screenshot shows the "Spawner Options" page of the Jupyter Hub. The header includes the Hogeschool Rotterdam logo and links for "Home" and "Token". The main content area has a heading "Spawner Options" and a sub-section "Welcome to Jupyter Hub on hr Research Drive Research Drive." It features a dropdown menu "Select your notebook" set to "DataScience Notebook" and an orange "Create your notebook" button.

*How to create a  
Proof of Concept [PoC]  
for Human Dialogues  
with large language Models*

*Definities / Omschrijvingen / Richtlijnen voor veilig gebruik*

“Generatieve AI”

# Generative AI according to TechTarget

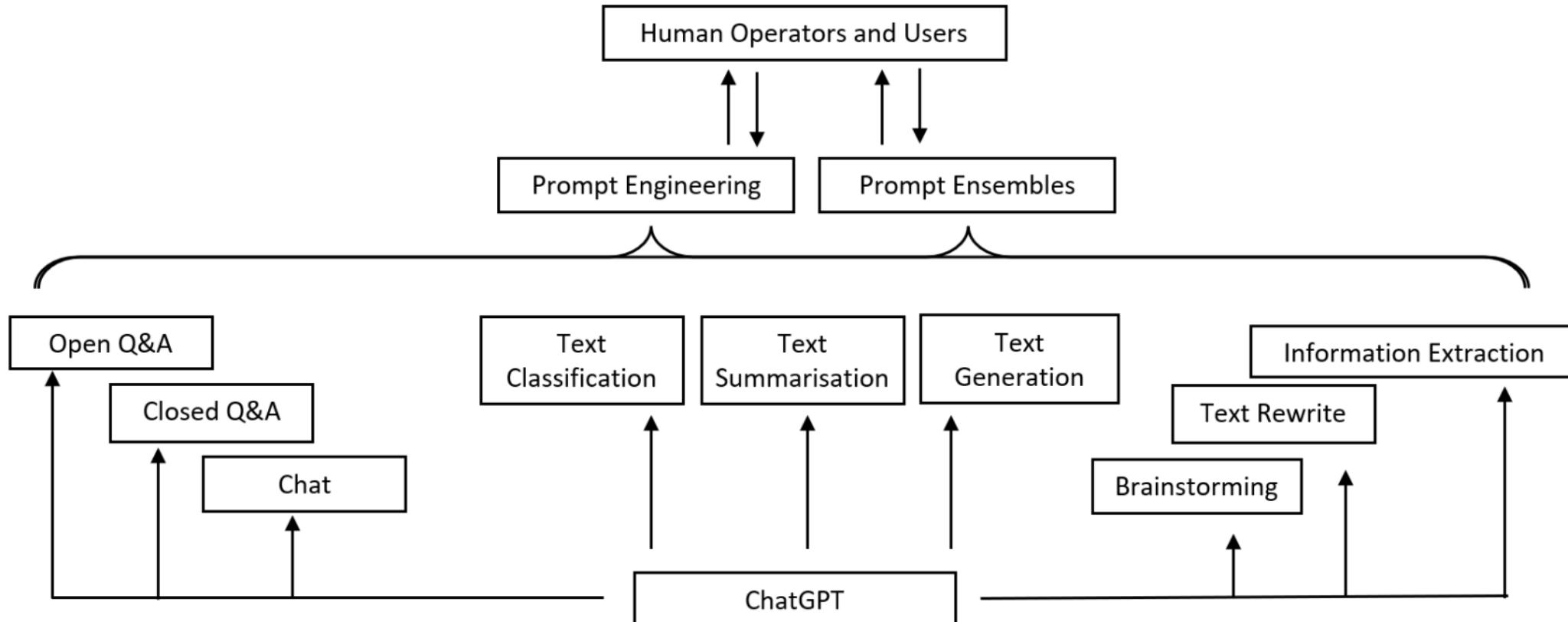
*Generative AI is a type of artificial intelligence technology that can produce various types of content, including text, imagery, audio and synthetic data.*

*The recent buzz around generative AI has been driven by the simplicity of new user interfaces for creating high-quality text, graphics and videos in a matter of seconds.*

*The technology, it should be noted, is not brand-new. Generative AI was introduced in the 1960s in chatbots.*

*But it was not until 2014, with the introduction of generative adversarial networks, or GANs --a type of machine learning algorithm-- that generative AI could create convincingly authentic images, videos and audio of real people.*

# ChatGPT: Use Cases



Conferences > 2023 IEEE International Confe... ⓘ

ChatGPT and Generative AI Guidelines for Addressing Academic Integrity and Augmenting Pre-Existing Chatbots

Publisher: IEEE

Cite This

PDF

Daswin De Silva ; Nishan Mills ; Mona El-Ayoubi ; Milos Manic ; Damminda Alahakoon All Authors

635  
Full  
Text Views





<https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/prompt-engineering>

Microsoft | Learn Documentation Training Credentials Q&A Code Samples Assessments Shows

Azure Product documentation ▾ Architecture ▾ Learn Azure ▾ Develop ▾ Resources ▾ Troubleshooting

Filter by title Learn / Azure / AI Services /

Azure OpenAI Service Documentation

Overview

- What is Azure OpenAI?
- Quotas and limits
- Models
- Legacy models
- Pricing ⓘ
- What's new
- Azure OpenAI FAQ

Quickstarts

Concepts

- Abuse monitoring
- Content filtering
- Embeddings
- Red teaming large language models (LLMs)

Intro to prompt engineering

Prompt engineering techniques

System message templates

Using your data (preview)

How-to

Tutorials

Responsible AI

Reference

Resources

# Introduction to prompt engineering

Article • 07/18/2023 • 2 contributors ⚙️ Feedback

In this article

Basics

Best practices

Space efficiency

Next steps

GPT-3, GPT-3.5, and GPT-4 models from OpenAI are prompt-based. With prompt-based models, the user interacts with the model by entering a text prompt, to which the model responds with a text completion. This completion is the model's continuation of the input text.

While these models are extremely powerful, their behavior is also very sensitive to the prompt. This makes prompt construction an important skill to develop.

Prompt construction can be difficult. In practice, the prompt acts to configure the model weights to complete the desired task, but it's more of an art than a science, often requiring experience and intuition to craft a successful prompt. The goal of this article is to help get you started with this learning process. It attempts to capture general concepts and patterns that apply to all GPT models. However it's important to understand that each model behaves differently, so the learnings may not apply equally to all models.

## Basics

This section covers the basic concepts and elements of GPT prompts.

Text prompts are how users interact with GPT models. As with all generative language models, GPT models attempt to produce the next series of words that are most likely to follow from the previous text. It's as if we're saying *What is the first thing that comes to your mind when I say <prompt>?*. The examples below demonstrate this behavior. Given the first words of famous content, the model is able to accurately continue the text.

Prompt	Completion
Four score and seven years ago our	fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal. [...]
"Call me Ishmael."	"Some years ago—never mind how long precisely—having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little and see the watery part of the world." [...]



HOGESCHOOL  
ROTTERDAM

<https://github.com/promtslab/Awesome-Prompt-Engineering>

Awesome-Prompt-Engineering Public

Watch 49 Fork 186 Star 2.3k

main 1 branch 0 tags Go to file Add file Code

monk1337 Merge pull request #9 from ArtificialZeng/main ... 45ddcda 2 weeks ago 116 commits

\_source Create data\_format.py 8 months ago

LICENSE Initial commit 8 months ago

README.md Merge pull request #9 from ArtificialZeng/main 2 weeks ago

code-of-conduct.md Create code-of-conduct.md 8 months ago

contributing.md Update contributing.md 8 months ago

README.md

## Awesome Prompt Engineering 🤖

This repository contains a hand-curated resources for Prompt Engineering with a focus on Generative Pre-trained Transformer (GPT), ChatGPT, PaLM etc

Prompt Engineering Course is coming soon..

awesome License Apache 2.0 PRs welcome Discord Community Open in Colab

## Table of Contents

- Papers
- Tools & Code
- APIs

### About

This repository contains a hand-curated resources for Prompt Engineering with a focus on Generative Pre-trained Transformer (GPT), ChatGPT, PaLM etc

[discord.gg/m88xfYMbK6](#)

machine-learning text-to-speech  
deep-learning prompt openai  
prompt-toolkit gpt text-to-image  
few-shot-learning text-to-video gpt-3  
prompt-learning prompt-tuning  
prompt-engineering prompt-generator  
promptengineering prompt-based-learning  
chatgpt chatgpt-api

Readme Apache-2.0 license  
Code of conduct  
Activity 2.3k stars  
49 watching 186 forks  
Report repository

### Releases

No releases published

### Packages

No packages published

### Contributors 11

# PROMPT-ENGINEERING

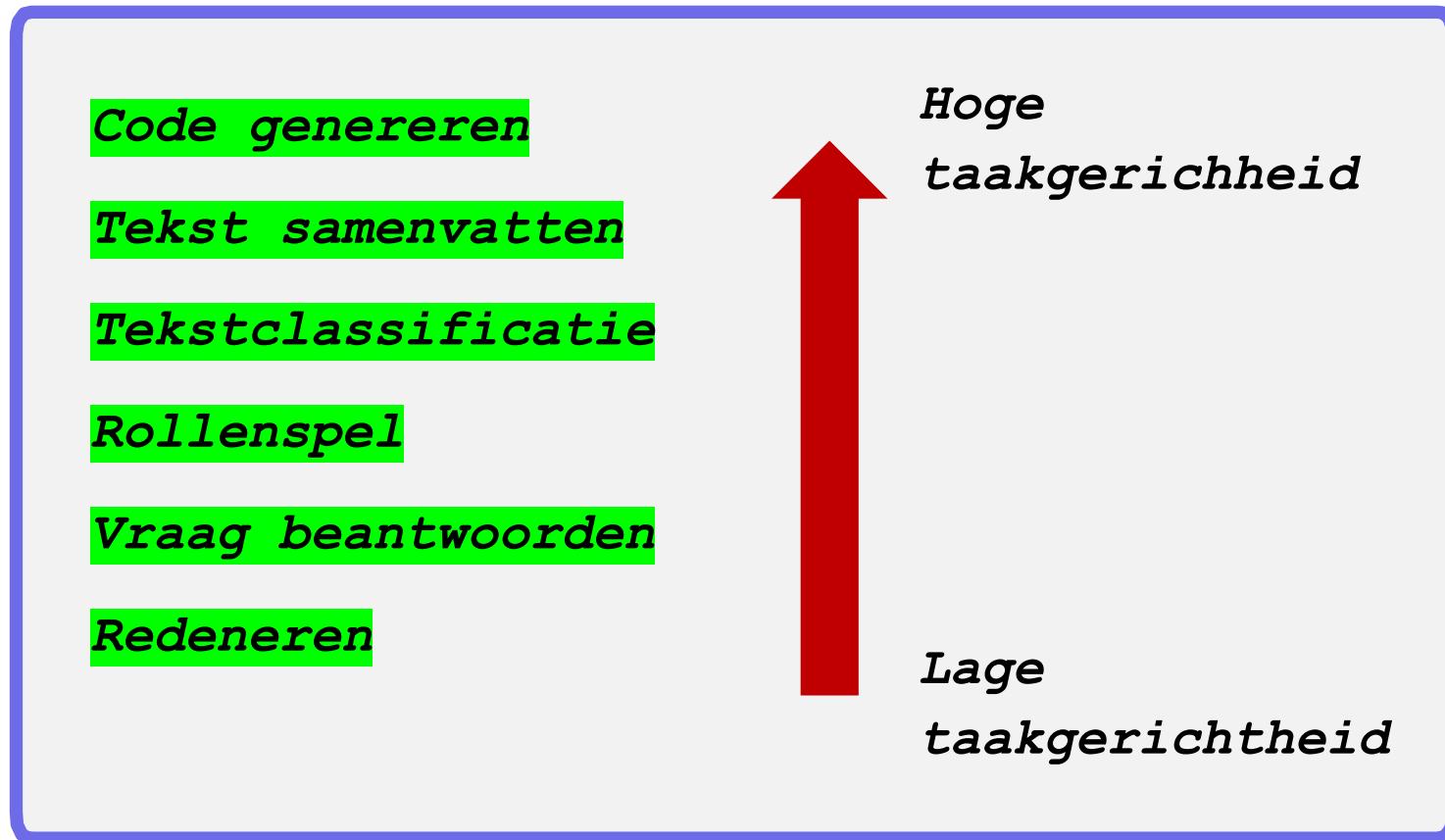
Het creatieve proces van het schrijven van een effectief ***prompt-recept*** wordt in het Engels "***prompt engineering***" genoemd.

Het schrijven van prompt-recepten

***---pseudo-Code---***

is een talige manier van het programmeren van "bevroren" voorgetraind taalmodellen.

# Taakgerichtheid van prompts



# Wat is het belang van Prompt Recepten Schrijven?

*Sturen van de mate van taakgerichtheid door  
reduceren van variatie in het antwoord zodat de  
kans groter wordt dat de uitkomst correct is.*

# Prompt Recept Structuur

Een prompt is opgebouwd uit de volgende elementen:

Instructie(s)

Context

Invoergegevens

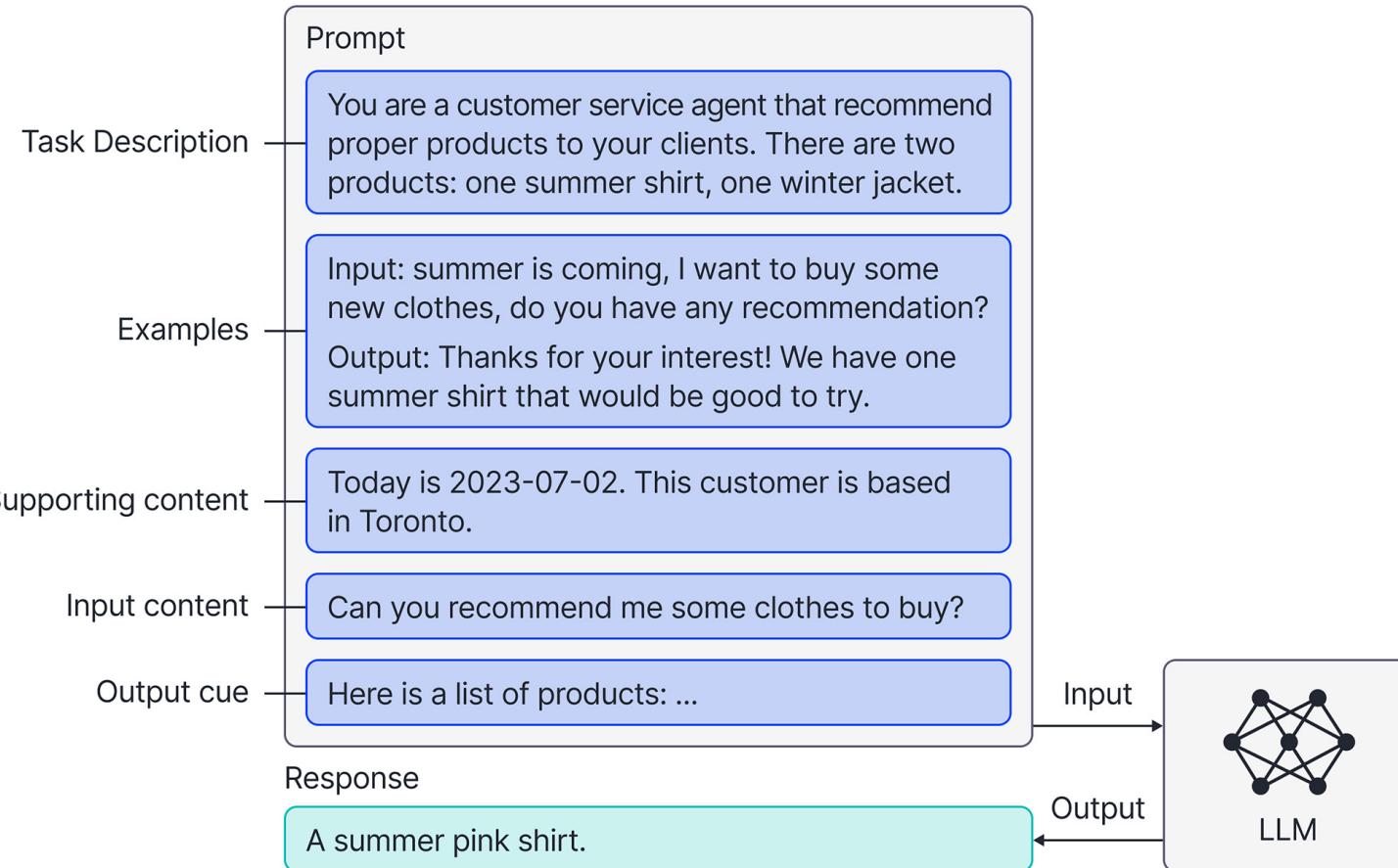
Uitvoer-indicator

*Classificeer de onderstaande tekst als neuraal, negatief of positief*

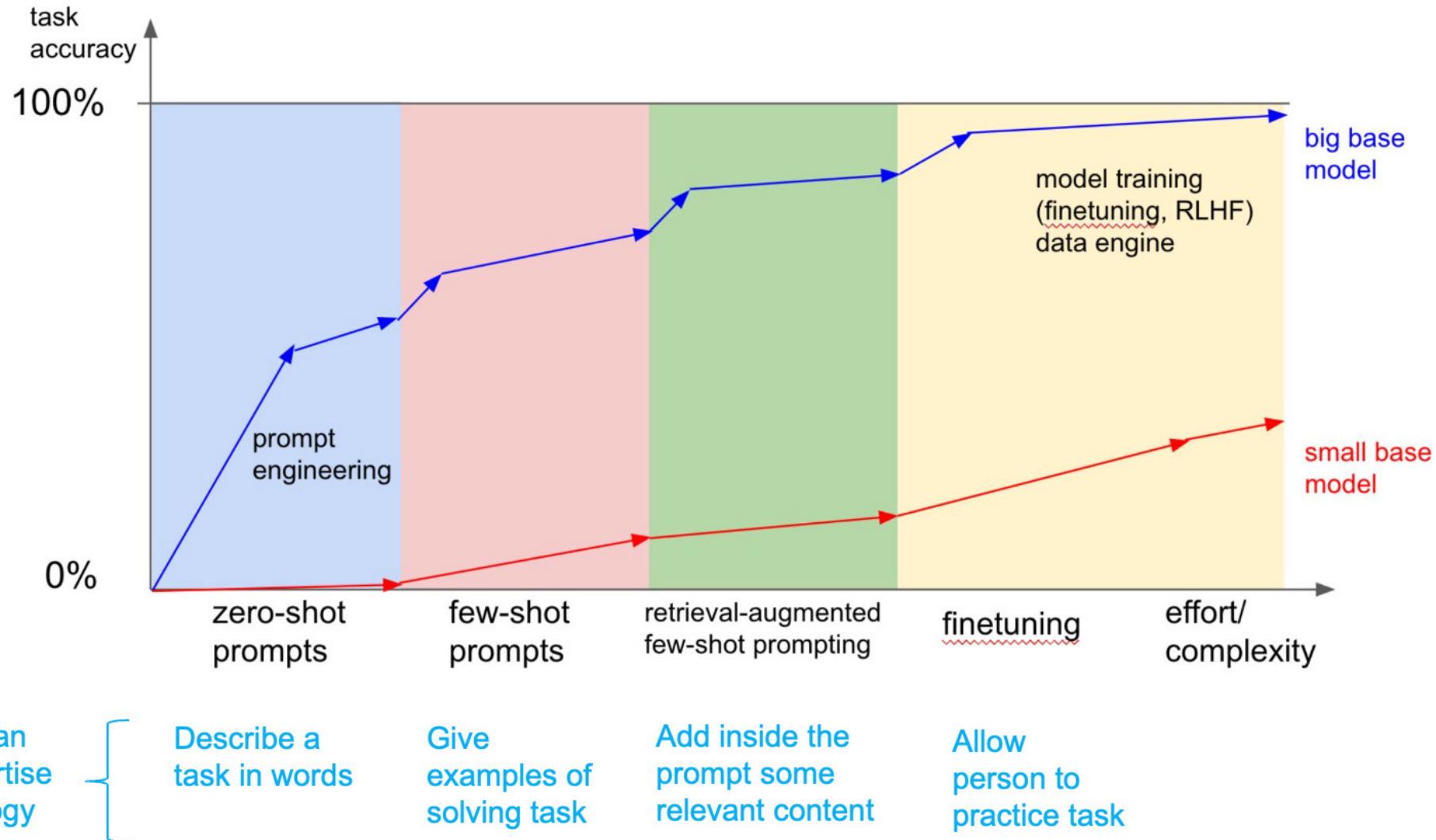
*Text: Ik vond het eten wel zoso.*

*Sentiment:*

# Prompt Recept Structuur



# Building a PoC dialogue



# *“digital Humans dialogues”*

*Hoe bouw  
en test je veilig  
Talige Generatieve  
AI-technologie?*

*Hoe bouw  
en test je veilig  
**Talige Generatieve  
AI-technologie?***

# Azure OpenAI service

OVERVIEW

## Build intelligent apps with AI models

Cutting edge models ▾

Quickly develop generative AI experiences with a diverse set of prebuilt and curated models from OpenAI, Meta and beyond.

[Try the Azure AI Studio](#)

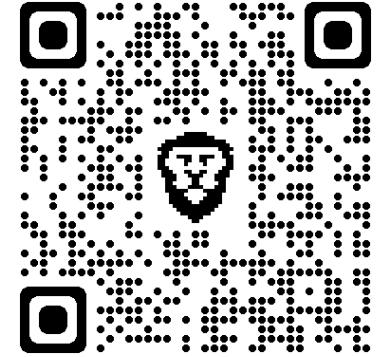
Data grounding ▾

Trust and transparency ▾

Data, privacy and security ▾

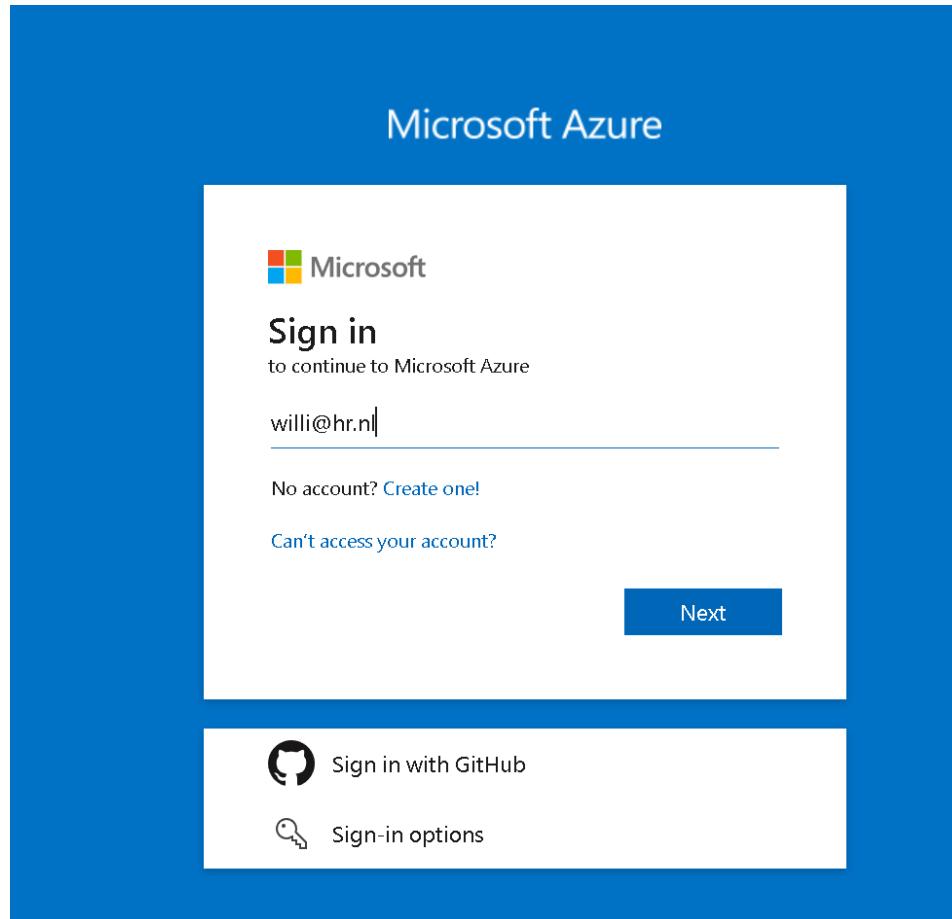
USE CASES

## Apply generative AI to a variety of use cases



<https://azure.microsoft.com/en-us/products/ai-services/openai-service>

portal.azure.com



# Log in

**Student number or personnel code**

---

**Password**

---

**Login**



HOGESCHOOL  
ROTTERDAM

### Azure services



### Resources

Recent   Favorite

Name	Type	Last Viewed
 GPT4-SWEDEN-GROUP	Azure OpenAI	6 days ago
 Azure for Students	Subscription	3 weeks ago
 CHATBOT02	Azure OpenAI	3 weeks ago
 DefaultResourceGroup-westeuropa	Resource group	2 months ago
 AV07	Speech service	7 months ago
 NLP	Resource group	7 months ago
 LLM01	Language understanding	7 months ago
 Visual Studio Professional Subscription	Subscription	7 months ago
 LLM01-Authoring	Language understanding	7 months ago
 WILLI107	Resource group	7 months ago
 cursusa1-900	Azure Machine Learning workspace	8 months ago
 cursusa9006361709869	Key vault	8 months ago

See all

### Navigate



Subscriptions



Resource groups



All resources



Dashboard

### Tools



Microsoft Learn   
Learn Azure with free online training from Microsoft



Azure Monitor  
Monitor your apps and infrastructure



Microsoft Defender for Cloud  
Secure your apps and infrastructure



Cost Management  
Analyze and optimize your cloud spend for free

### Useful links

### Azure mobile app



HOGESCHOOL  
ROTTERDAM

## Create Azure OpenAI ...

## Azure services

[Create a resource](#)[Azure OpenAI](#)[Subscriptions](#)[Resource groups](#)[Education](#)[Cost Management](#)[All resources](#)[Azure AI services](#)[App Services](#)[More services](#)

## Resources

[Recent](#) [Favorite](#)

Name	Type	Last Viewed
GPT4-SWEDEN-GROUP	Azure OpenAI	8 minutes ago
cursusai9006361709869	Key vault	9 minutes ago
Azure for Students	Subscription	9 minutes ago
Visual Studio Professional Subscription	Subscription	10 minutes ago
NLP	Resource group	11 minutes ago
CHATBOT02	Azure OpenAI	3 weeks ago
DefaultResourceGroup-westeurope	Resource group	2 months ago
AV07	Speech service	7 months ago
LLM01	Language understanding	7 months ago
LLM01-Authoring	Language understanding	7 months ago
WILLI107	Resource group	7 months ago
cursusai-900	Azure Machine Learning workspace	8 months ago

[See all](#)[Review the Azure OpenAI code of conduct](#)

# Create and deploy an Azure OpenAI Service resource

Article • 09/06/2023 • 4 contributors

Feedback

Choose your preferred resource creation method

Portal **CLI** PowerShell

## In this article

[Prerequisites](#)[Create a resource](#)[Deploy a model](#)[Next steps](#)

This article describes how to get started with Azure OpenAI Service and provides step-by-step instructions to create a resource and deploy a model. You can create resources in Azure in several different ways:

- The [Azure portal](#)
- The REST APIs, the Azure CLI, PowerShell, or client libraries
- Azure Resource Manager (ARM) templates

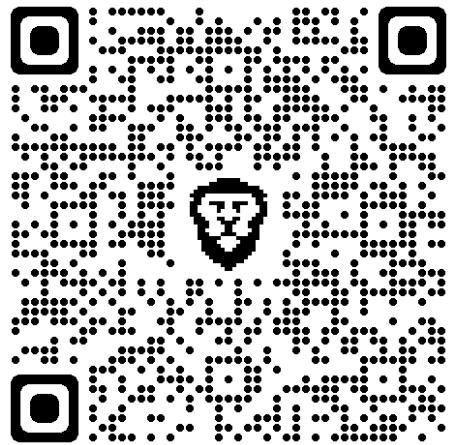
In this article, you review examples for creating and deploying resources in the Azure portal and with the Azure CLI.

## Prerequisites

- An Azure subscription. [Create one for free](#).
- Access granted to Azure OpenAI in the desired Azure subscription.
- Access permissions to [create Azure OpenAI resources](#) and to [deploy models](#).

 **Note**

Currently, you must submit an application to access Azure OpenAI Service. To apply for access, complete [this form](#). If you need assistance, open an issue on this repository to contact Microsoft.



# Request Access to Azure OpenAI Service

\* Required

## Please read all instructions carefully and complete form as instructed

Thank you for your interest in Azure OpenAI Service. Please submit this form to register for approval to access and use Azure OpenAI's Limited Access text and code and/or DALL-E 2 text to image models (as indicated in the form). All use cases must be registered. Azure OpenAI Service requires registration and is currently only available to approved enterprise customers and partners. Learn more about limited access to Azure OpenAI Service [here](#).

**Limited access scenarios:** When evaluating which scenarios to onboard, we consider who will directly interact with the application, who will see the output of the application, whether the application will be used in a high-stakes domain (e.g., medical), and the extent to which the application's capabilities are tightly scoped. In general, applications in high stakes domains will require additional mitigations and are more likely to be approved for applications with internal-only users and internal-only audiences. Applications with broad possible uses, including content generation capabilities, are more likely to be approved if 1) the domain is not high stakes and users are authenticated or 2) in the case of high stakes domains, anyone who views or interacts with the content is internal to your company.

Please be sure to visit the [Azure OpenAI Service's transparency note](#), which provides information and guidelines for responsible use of the service as well as system limitations that may be applicable to your scenario.

If you are a **current Azure OpenAI customer** and would like to add additional use cases, please fill out the [Azure OpenAI Additional Use Case form](#)



[https://customervoice.microsoft.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180B\\_HbR7en2Ais5pxKtso\\_Pz4b1\\_xUOFA5Qk1UWDRBMjg0WFhPMkIzTzhKQ1dWNyQIQCN0PWcu](https://customervoice.microsoft.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180B_HbR7en2Ais5pxKtso_Pz4b1_xUOFA5Qk1UWDRBMjg0WFhPMkIzTzhKQ1dWNyQIQCN0PWcu)

Azure AI | Azure OpenAI Studio

« Azure AI Studio > Chat playground

## Chat playground

**Assistant setup**

System message  Add your data (preview)

Save changes

**Specify how the chat should act**

Use a template to get started, or just start writing your own system message below. Want some tips? [Learn more](#)

**Use a system message template**

Select a template

**System message ⓘ**

You are an AI assistant that helps people find information.

**Examples ⓘ**

+ Add an example

**Sample Code**

You can use the following code to start integrating your current prompt and settings into your application

<https://gpt4-sweden-group.openai.azure.com/>  python

```
1 #Note: The openai-python library support for Azure OpenAI is in preview.
2 import os
3 import openai
4 openai.api_type = "azure"
5 openai.api_base = "https://gpt4-sweden-group.openai.azure.com/"
6 openai.api_version = "2023-07-01-preview"
7 openai.api_key = os.getenv("OPENAI_API_KEY")
8
9 response = openai.chatcompletion.create(
10     engine="GPT4-32K",
11     messages = [{"role": "system", "content": "You are an AI
assistant that helps people find information."},
12 {"role": "user", "content": "A neutron star is the collapsed core of a
massive supergiant star, which had a total mass of between 10 and 25
solar masses, possibly more if the star was especially metal-rich.
Neutron stars are the smallest and densest stellar objects, excluding
black holes and hypothetical white holes, quark stars, and strange
stars. Neutron stars have a radius on the order of 10 kilometres (6.2
mi) and a mass of about 1.4 solar masses. They result from the
supernova explosion of a massive star, combined with gravitational
collapse, that compresses the core past white dwarf star density to
that of atomic nuclei.\n\nQ: How are neutron stars created?\nA:"}],
13     "role": "assistant", "content": "Neutron stars are created from the
supernova explosion of a massive star, combined with gravitational
collapse, that compresses the core past white dwarf star density to
that of atomic nuclei.\n\nQ: How are neutron stars created?\nA:"}]
```

**Endpoint ⓘ**

<https://gpt4-sweden-group.openai.azure.com/openai/deployments/GPT4-3...>

**Key ⓘ**

.....

You should use environment variables or a secret management tool like Azure Key Vault to prevent accidental exposure of your key in applications. [Learn more](#)

Copy  Close

## Azure chat completions example (preview)

In this example we'll try to go over all operations needed to get chat completions working using the Azure endpoints.

This example focuses on chat completions but also touches on some other operations that are also available using the API. This example is meant to be a quick way of showing simple operations and is not meant as a tutorial.

```
 1 import os
 2 import openai
 3 openai.api_type = "azure"
 4 openai.api_base = "https://gpt4-sweden-group.openai.azure.com/"
 5 openai.api_version = "2023-07-01-preview"
 6 openai.api_key = "ded218c778894f6da4d3c595c6904194"
 7
 8
 9 #!setx AZURE_OPENAI_KEY "ded218c778894f6da4d3c595c6904194"
10 #!setx AZURE_OPENAI_ENDPOINT "https://gpt4-sweden-group.openai.azure.com/openai/deployments/gpt4-32k/chat/completions?api-version=2023-07-01-preview"
11
12 response = openai.chatCompletion.create(
13     engine="GPT4-32K",
14     messages = [
15         {"role": "system", "content": "You are a helpful assistant."},
16         {"role": "user", "content": "Does Azure OpenAI support customer managed keys?"},
17         {"role": "assistant", "content": "Yes, customer managed keys are supported by Azure OpenAI."},
18         {"role": "user", "content": "Do other Azure AI services support this too?"}
19     ],
20     temperature=0.7,
21     max_tokens=800,
22     top_p=0.95,
23     frequency_penalty=0,
24     presence_penalty=0,
25     stop=None)
26
27
28 print(response)
29 print(response['choices'][0]['message']['content'])
30
```

✓ 17.9s

```
{
  "id": "chatcmpl-8Cp65kWmkKF58MB1TtRnUPd60Ak",
  "object": "chat.completion",
  "created": 1698067083,
  "model": "gpt-4-32k",
  "prompt_filter_results": [
    {
      "prompt_index": 0,
      "content_filter_results": {
        "hate": {
          "filtered": false,
          "severity": "safe"
        },
        "self_harm": {
          "filtered": false,
          "severity": "safe"
        },
        "sexual": {
          "filtered": false,
          "severity": "safe"
        },
        "violence": {
          "filtered": false,
          "severity": "safe"
        }
      }
    }
  ]
}
```

PROBLEMS 797 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER CODEWHISPERER REFERENCE LOG

S C:\Users\rob>



HOGESCHOOL  
ROTTERDAM

Learn / Azure / AI Services /

# Learn how to generate or manipulate text

Article • 08/17/2023 • 2 contributors

In this article

- Design prompts
- Classify text
- Trigger ideas
- Conduct conversations

Show 5 more

Azure OpenAI Service provides a **completion endpoint** that can be used for a wide variety of tasks. The endpoint supplies a simple yet powerful text-in, text-out interface to any [Azure OpenAI model](#). To trigger the completion, you input some text as a prompt. The model generates the completion and attempts to match your context or pattern. Suppose you provide the prompt "As Descartes said, I think, therefore" to the API. For this prompt, Azure OpenAI returns the completion endpoint "I am" with high probability.

The best way to start exploring completions is through the playground in [Azure OpenAI Studio](#). It's a simple text box where you enter a prompt to generate a completion. You can start with a simple prompt like this one:

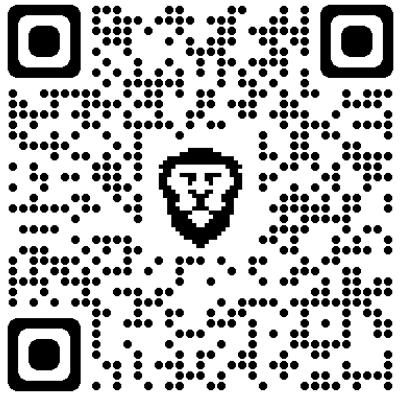
Console Copy

```
write a tagline for an ice cream shop
```

After you enter your prompt, Azure OpenAI displays the completion:

Console Copy

```
we serve up smiles with every scoop!
```



<https://learn.microsoft.com/en-us/azure/ai-services/openai/how-to/completions>

## GPT-3.5 models

GPT-3.5 Turbo is used with the Chat Completion API. GPT-3.5 Turbo (0301) can also be used with the Completions API. GPT3.5 Turbo (0613) only supports the Chat Completions API.

GPT-3.5 Turbo version 0301 is the first version of the model released. Version 0613 is the second version of the model and adds function calling support.

Model ID	Base model Regions	Fine-Tuning Regions	Max Request (tokens)	Training Data (up to)
<code>gpt-35-turbo<sup>1</sup></code> (0301)	East US, France Central, South Central US, UK South, West Europe	N/A	4,096	Sep 2021
<code>gpt-35-turbo</code> (0613)	Australia East, Canada East, East US, East US 2, France Central, Japan East, North Central US, Sweden Central, Switzerland North, UK South	North Central US, Sweden Central	4,096	Sep 2021
<code>gpt-35-turbo-16k</code> (0613)	Australia East, Canada East, East US, East US 2, France Central, Japan East, North Central US, Sweden Central, Switzerland North, UK South	N/A	16,384	Sep 2021
<code>gpt-35-turbo-instruct</code> (0914)	East US, Sweden Central	N/A	4,097	Sep 2021

<sup>1</sup> Version `0301` of gpt-35-turbo will be retired no earlier than July 5, 2024. See [model updates](#) for model upgrade behavior.



# {Knowledge Dissemination & Curation}

High quality Dutch reviews on state-of-the-art AI



De (on)mogelijkheden van kunstmatige intelligentie in het onderwijs



In opdracht van:  
Ministerie van Onderwijs, Cultuur & Wetenschap

Project:  
2018.06.06

Publicatienummer:  
2018.06.1828 v1.0.116

Datum:  
Utrecht, 21 januari 2019

Auteurs:  
ir. Tommy van der Vorst  
ir. Nick Jelicic  
mr. Marc de Vries  
Julie Albers

# {Knowledge Dissemination & Curation}



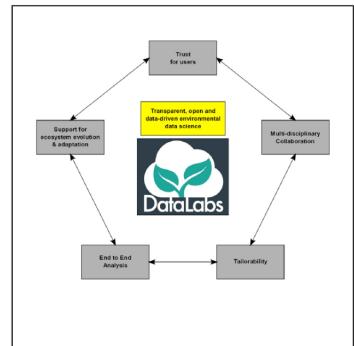
HOGESCHOOL  
ROTTERDAM

## Learn from prime examples

### Patterns

#### Tackling the Challenges of 21<sup>st</sup>-Century Open Science and Beyond: A Data Science Lab Approach

##### Graphical Abstract



##### Authors

Michael J. Hollaway, Graham Dean, Gordon S. Blair, Mike Brown, Peter A. Henrys, John Watkins

##### Correspondence

mjholloway@ceh.ac.uk

##### In Brief

We present the "data science lab" concept as a potential solution to champion cross-disciplinary and open science. Data science labs are cloud-based, collaborative, and tailorable platforms enabling users with different requirements and expertise to find data-driven solutions to a wide range of environmental challenges. We present examples of methodological and infrastructural developments using data science labs along with a detailed research roadmap to serve as a focal point for developing a more data-driven and transparent approach to environmental data science.

##### Highlights

- Offer a vision of data science labs as open, collaborative platforms in the cloud
- Discussion of how data science labs support open and transparent science
- Discussion of experiences around implementing data labs in practice
- The definition of a roadmap of research challenges around virtual data labs

Hollaway et al., 2020, Patterns 1, 100103  
October 9, 2020 © 2020 The Author(s).  
<https://doi.org/10.1016/j.patter.2020.100103>

CellPress

[Tackling the Challenges of 21st-Century Open Science and Beyond: A Data Science Lab Approach - ScienceDirect](#)

### THEME ARTICLE: JUPYTER IN COMPUTATIONAL SCIENCE

#### Using Jupyter for Reproducible Scientific Workflows

Marijan Beg Faculty of Engineering and Physical Sciences, University of Southampton, SO17 1BJ, Southampton, U.K.

Juliette Taka, Logilab, 75013, Paris, France

Thomas Kluyver European XFEL, 22869, Schenefeld, Germany

Alexander Konvalinov School of Computer Science, University of St. Andrews, KY16 9SX, St. Andrews, U.K.

Min Ragan-Kelley, Simula Research Laboratory, 1364, Fornebu, Norway

Nicolas M. Thiéry Laboratoire de Recherche en Informatique, Université Paris-Saclay, CNRS, 91405, Orsay, France

Hans Fangohr Max Planck Institute for the Structure and Dynamics of Matter, 22761, Hamburg, Germany

Literate computing has emerged as an important tool for computational studies and open science, with growing folklore of best practices. In this work, we report two case studies—one in computational magnetism and another in computational mathematics—where domain-specific software was exposed to the Jupyter environment. This enables high level control of simulations and computation, interactive exploration of computational results, batch processing on HPC resources, and reproducible workflow documentation in Jupyter notebooks. In the first study, *Ubertmag* drives existing computational micromagnetics software through a domain-specific language embedded in Python. In the second study, a dedicated Jupyter kernel interfaces with the GAP system for computational discrete algebra and its dedicated programming language. In light of these case studies, we discuss the benefits of this approach, including progress toward more reproducible and reusable research results and outputs, notably through the use of infrastructure such as *JupyterHub* and *Binder*.

Research usually results in a publication that presents and shares the obtained findings and conclusions. For a publication to be scientifically valid, it must present the methodology rigorously, so that readers can follow the "recipe" and reproduce the results. If this criterion is met, the publication is considered reproducible. *Reproducible* publications are more easily reusable and, thus, provide a significant opportunity to make (often tax-payer funded) research more impactful. However, the reproducibility of computational work is usually hindered not only by a lack of data or metadata but also by a lack of details on the procedure and tools used.

1521-9655 © 2021 IEEE  
Digital Object Identifier 10.1109/MCSE.2021.3052101  
Date of publication 15 January 2021; date of current version 25 March 2021.

36

Computing in Science & Engineering

Published by the IEEE Computer Society

March/April 2021

### Patterns

CellPress  
OPEN ACCESS

#### Tutorial

#### Ready, Steady, Go AI: A practical tutorial on fundamentals of artificial intelligence and its applications in phenomics image analysis

Farid Nakhle<sup>1</sup> and Antoine L. Harfouche<sup>1,\*</sup>

<sup>1</sup>Department for Innovation in Biological, Agro-food and Forest systems, University of Tuscia, Via S. Camillo de Lellis, Viterbo 01100, Italy

\*Correspondence: [alharfouche@unitus.it](mailto:alharfouche@unitus.it)

<https://doi.org/10.1016/j.patter.2021.100323>

**THE BIGGER PICTURE** Advances in AI technologies have the potential to significantly increase our ability to turn plant phenomics data into valuable insights. However, performing such analyses requires specialized programming skills commonly reserved for computer scientists. We created an interactive tutorial with free, open-source, and FAIR notebooks that can aid researchers to conduct such analyses without the need for an extensive coding experience. We supplemented it with a practical guide on how to implement AI and X-AI algorithms that augment and complement human experience in classifying tomato leaf diseases and spider mites. Our tutorial is not only applicable to other stresses but also transferable to other plants and research domains, making it possible for researchers from various scientific fields to generate insights into their data. We expect our notebooks to be of high interest to those who want to enhance the performance and sustainability of our agricultural systems through phenomics.

Production Data science output is validated, understood, and regularly used for multiple domains/platforms

#### SUMMARY

High-throughput image-based technologies are now widely used in the rapidly developing field of digital phenomics and are generating ever-increasing amounts and diversity of data. Artificial intelligence (AI) is becoming a game changer in turning the vast seas of data into valuable predictions and insights. However, this requires specialized programming skills and an in-depth understanding of machine learning, deep learning, and ensemble learning algorithms. Here, we attempt to methodically review the usage of different tools, technologies, and services available to the phenomics data community and show how they can be applied to selected problems in explainable AI-based image analysis. This tutorial provides practical and useful resources for novices and experts to harness the potential of the phenomic data in explainable AI-led breeding programs.

#### HELLO, WORLD!

In his tutorial memoranda introducing B and C programming languages in 1972 and 1975, respectively, Brian Kernighan wrote an example code that prints "hello, world!" on the computer terminal.<sup>1,2</sup> In 1978, he reused it in "The C Programming Language" book, a worldwide bestseller, which he co-authored with Dennis Ritchie, who originally designed and implemented the language.<sup>3</sup> Ever since, "hello, world!" has become the first program taught when learning a programming language with a purpose to illustrate its basic syntax and verify that the programming environment is properly set up. Seeing these two words on the screen means the code can compile and run correctly.

While this example is used as an illustration for almost every programming language introduced after C, it is also being incorporated in tutorials for older programming languages, such as list processor (LISP), the first high-level functional programming language, designed by John McCarthy in 1958.<sup>4</sup> LISP became predominant over the information processing language (IPL), the first programming language tailored for artificial intelligence (AI) programming, released in the style of low-level assembly language in 1956.<sup>5,6</sup>

Although IPL was introduced in 1956, the roots of AI date back to 1947 when Alan Turing delivered what is perhaps the earliest public lecture on computer intelligence.<sup>7</sup> In 1950, his paper discussed how to build intelligent machines and test their intelligence.<sup>8</sup> Five years later, Allen Newell, Clifford Shaw, and Herbert

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

[Using Jupyter for Reproducible Scientific Workflows | IEEE Journals & Magazine | IEEE Xplore](#)

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8441561/>

# {Knowledge Dissemination & Curation}



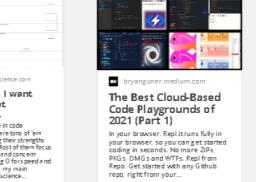
HOGESCHOOL  
ROTTERDAM

 newsmap.js.herokuapp.com

## NewsMap.js

A squarified treemap visualization of Google News based on the original newsmap.jp

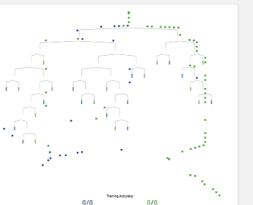


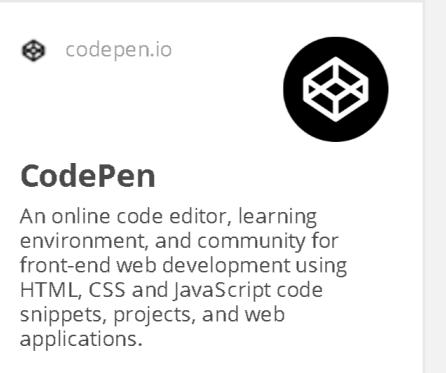


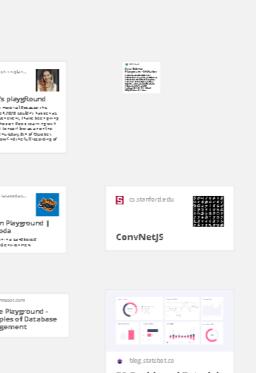


**Making predictions**

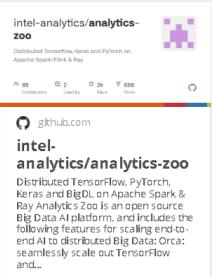
The newly-trained decision tree model determines whether a home is in San Francisco or New York by running each point through the branches.

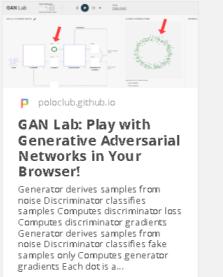




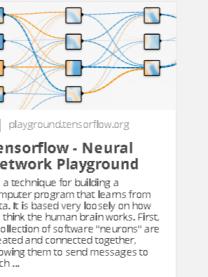








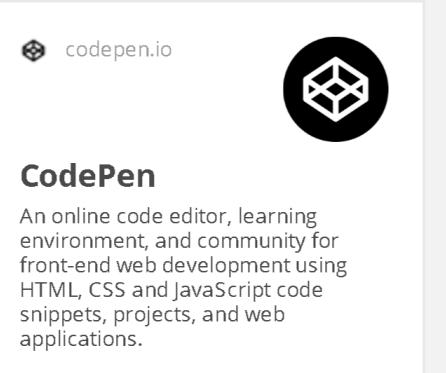


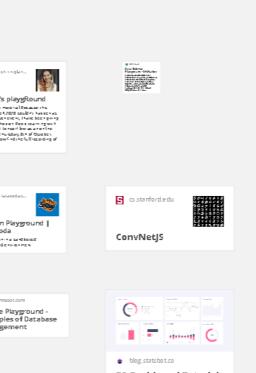




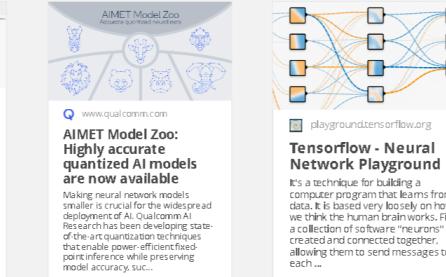


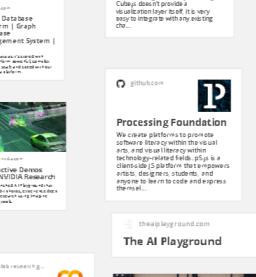




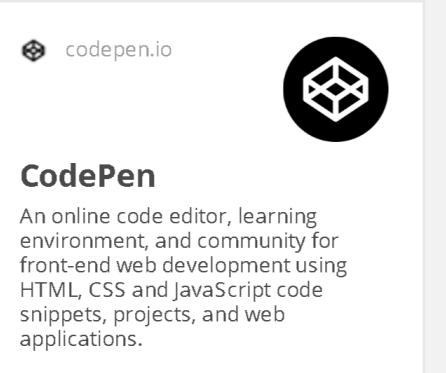


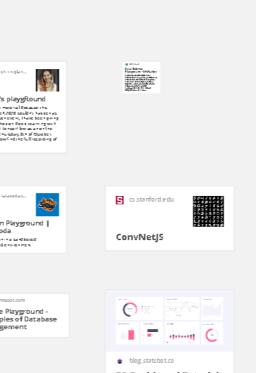




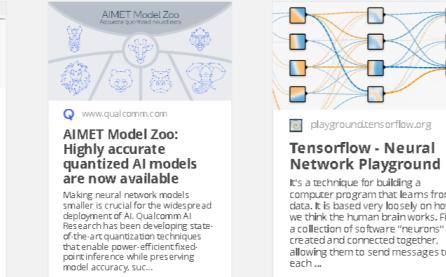


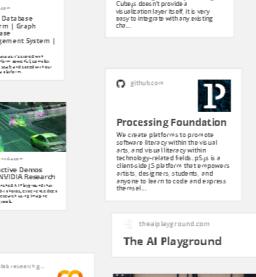


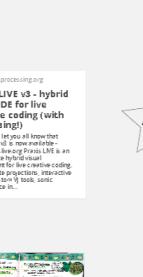


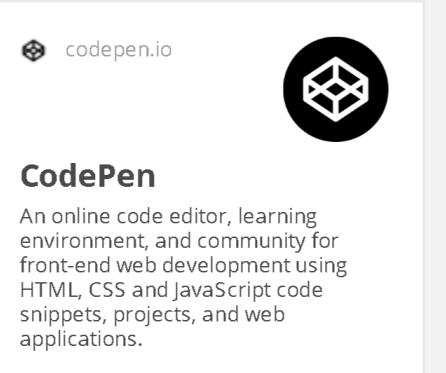


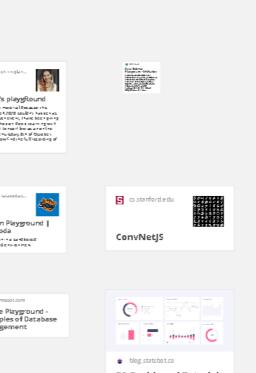




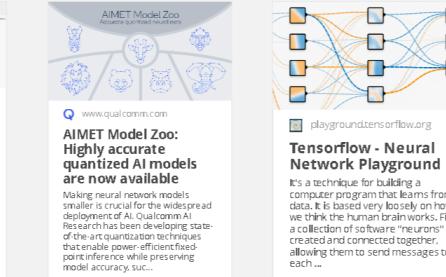


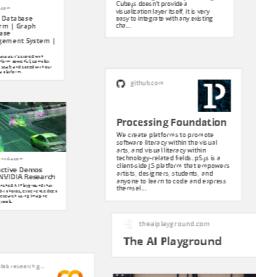




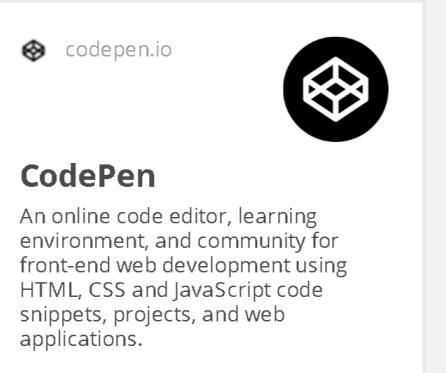


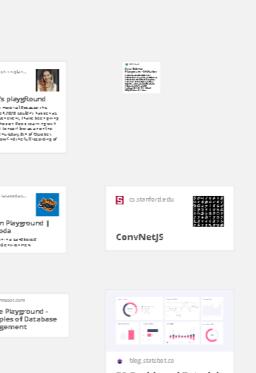




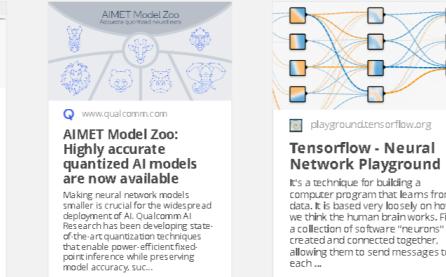


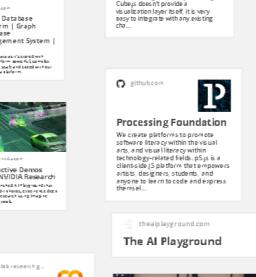




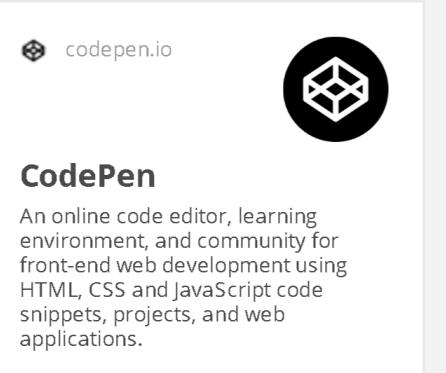


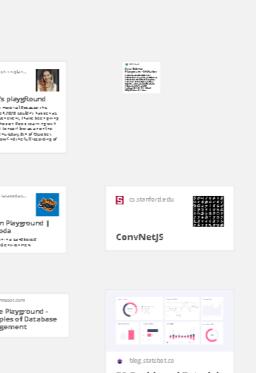




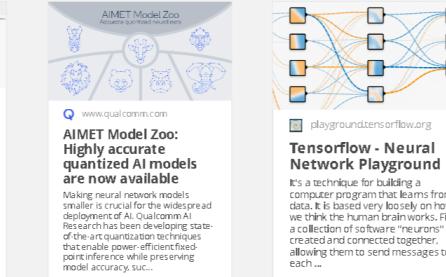


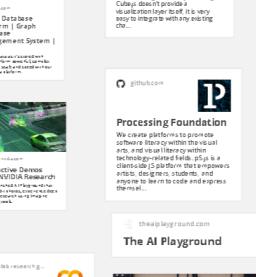




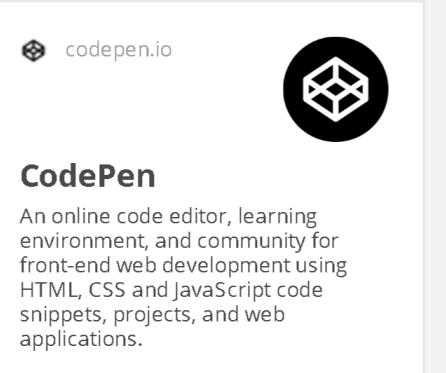


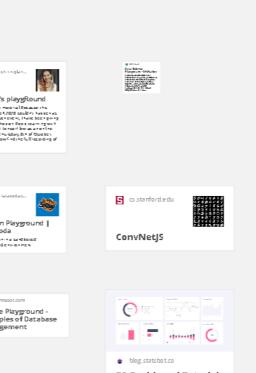




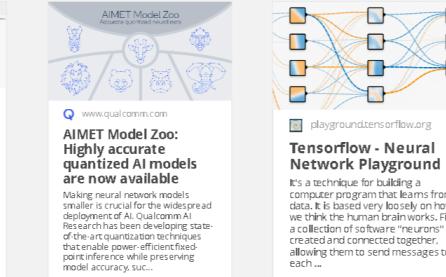


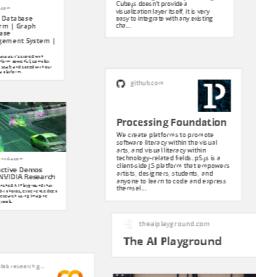


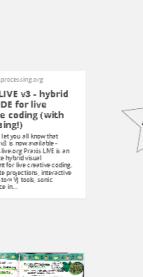


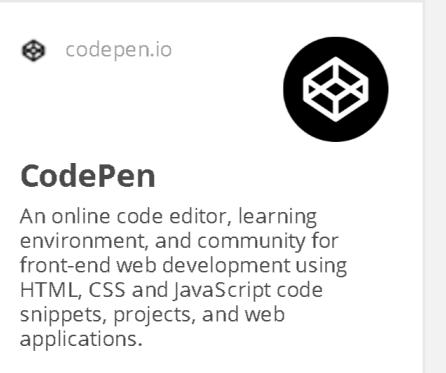


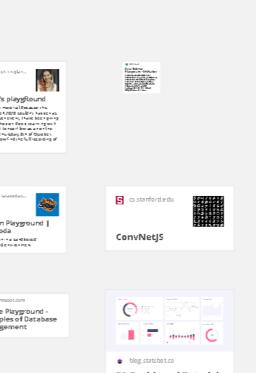




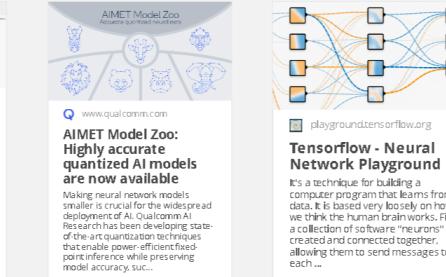


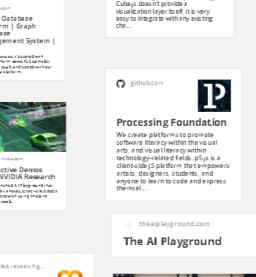




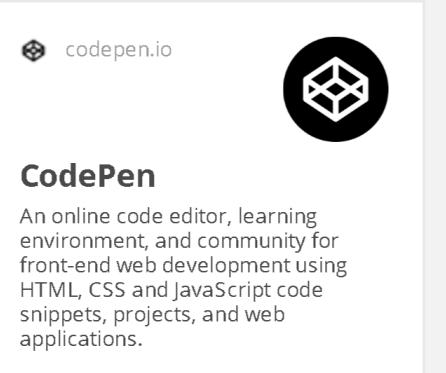


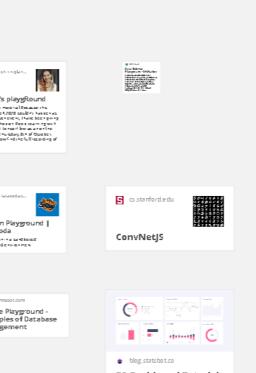




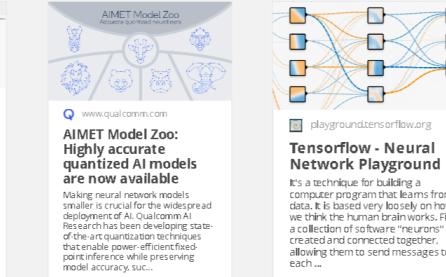


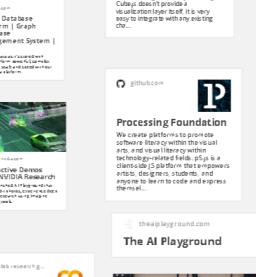




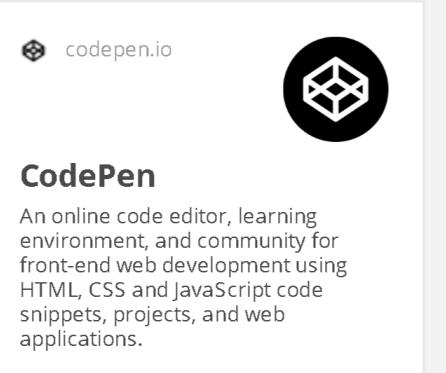


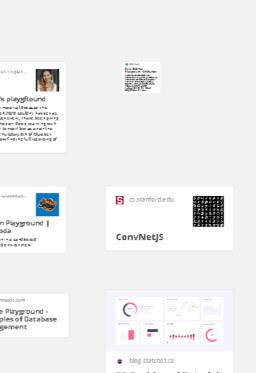




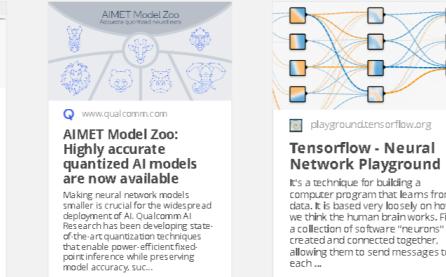


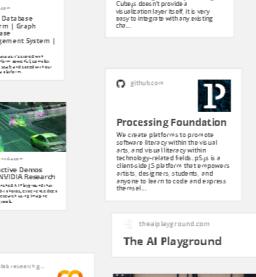




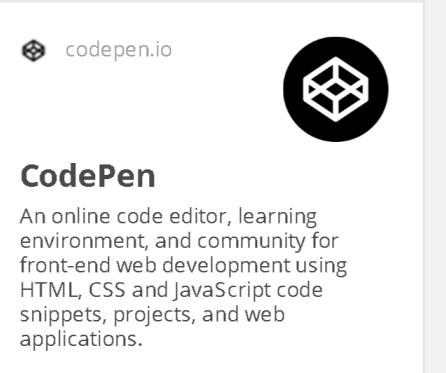


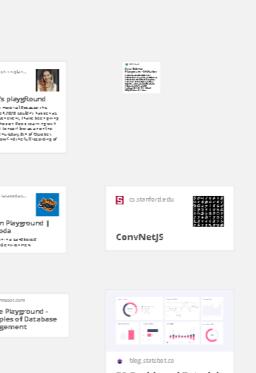




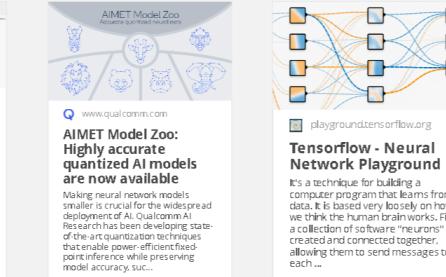


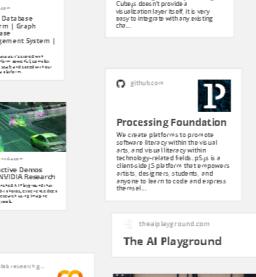




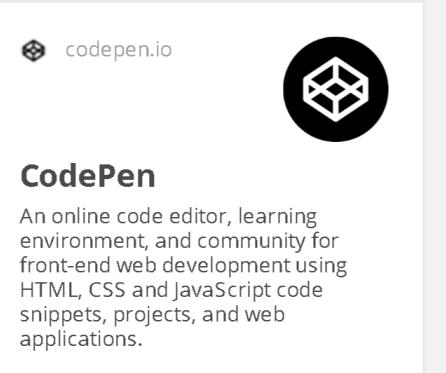


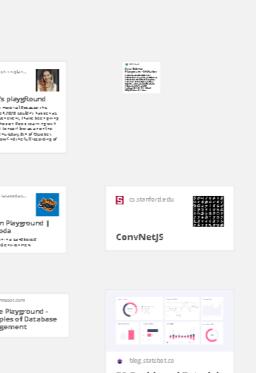




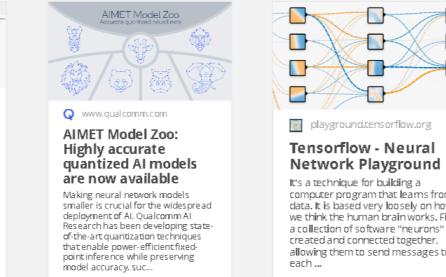


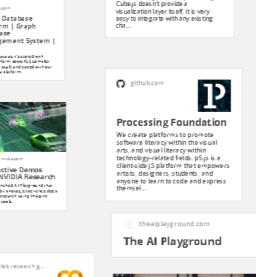




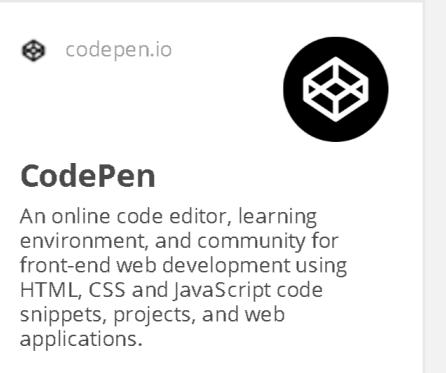


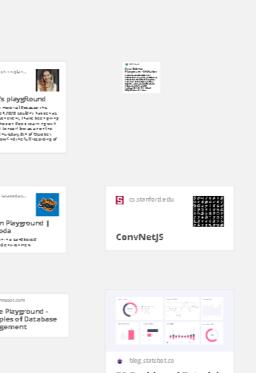




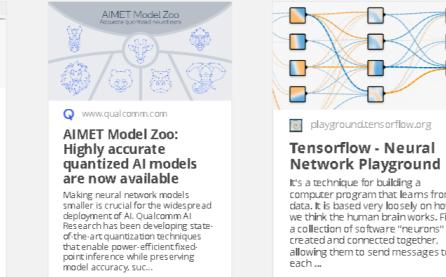


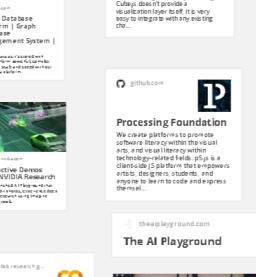




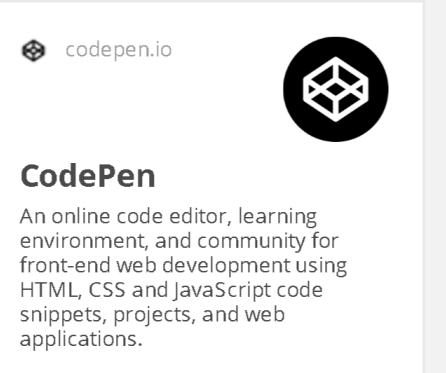


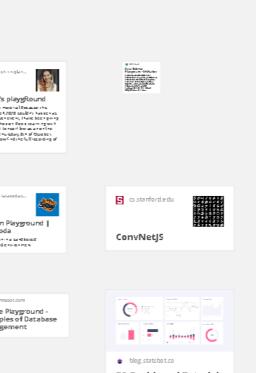




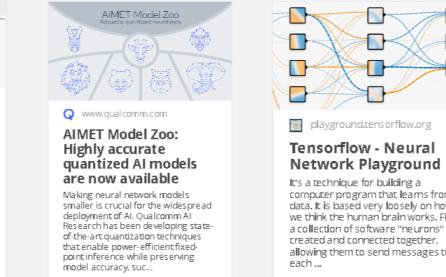


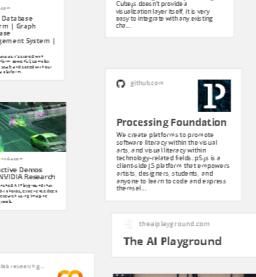




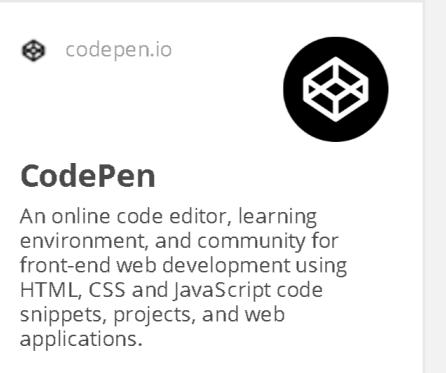


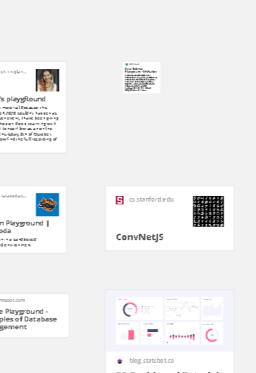




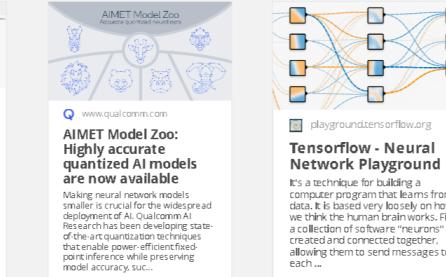


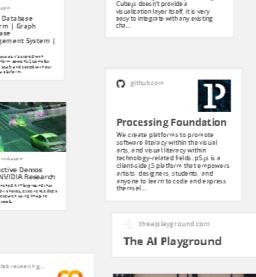


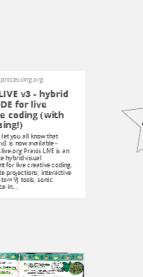


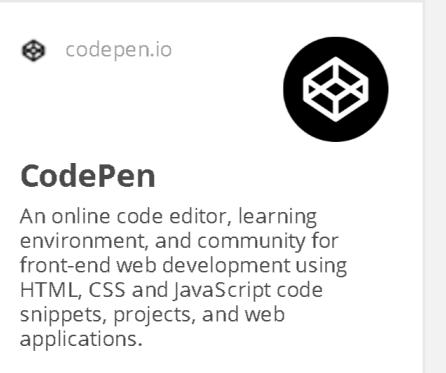


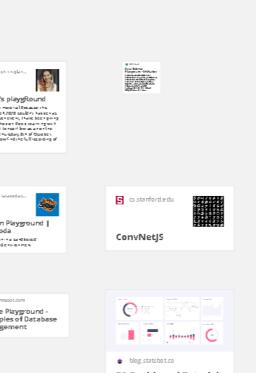




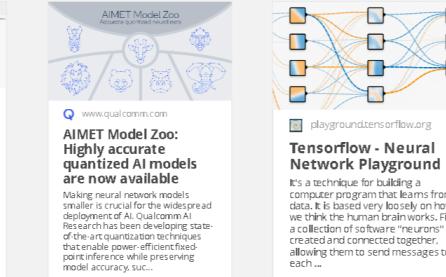


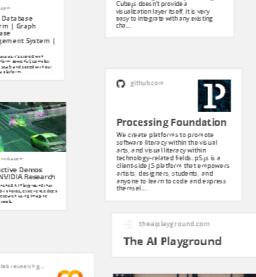




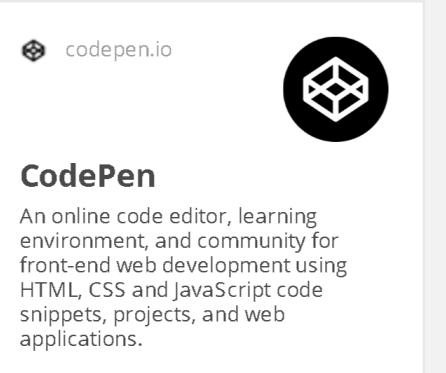


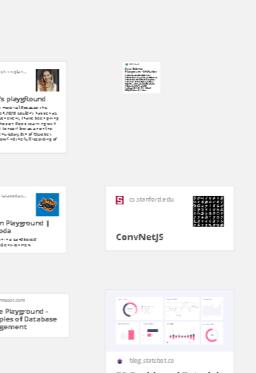




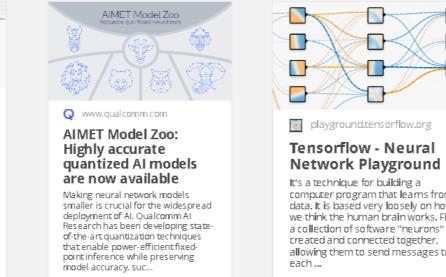


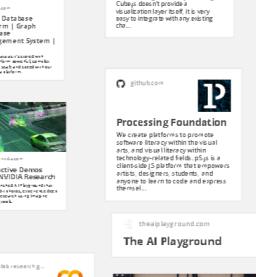




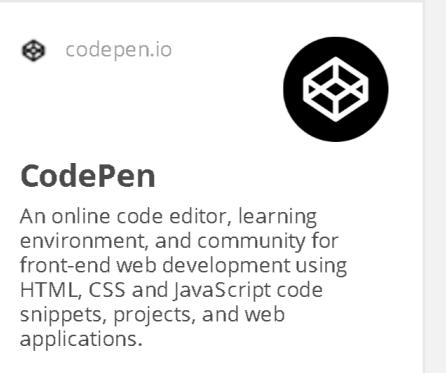


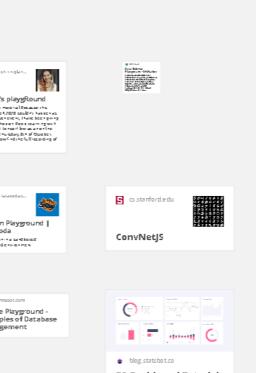




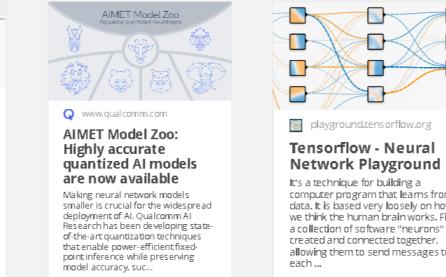


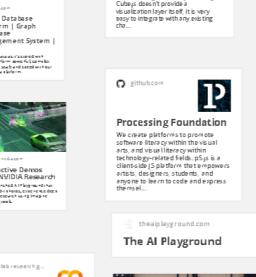




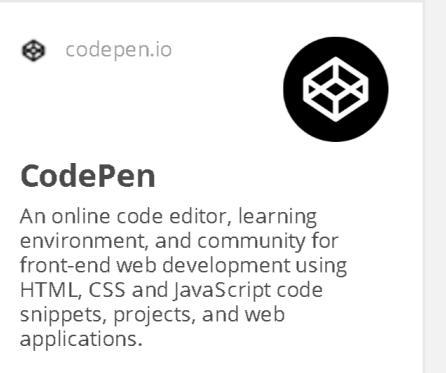


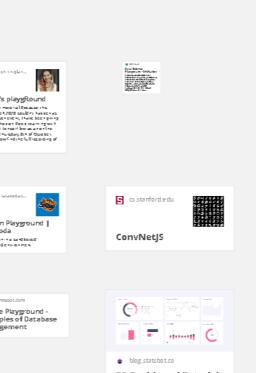




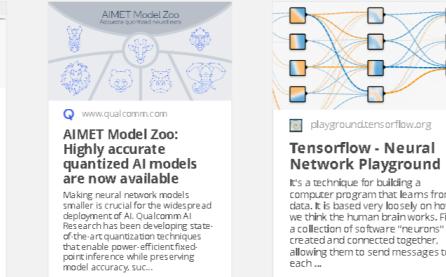


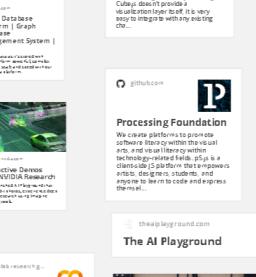




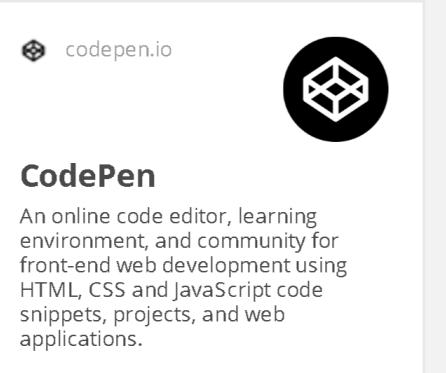


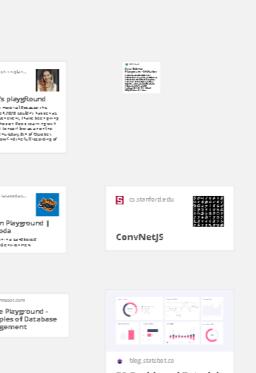




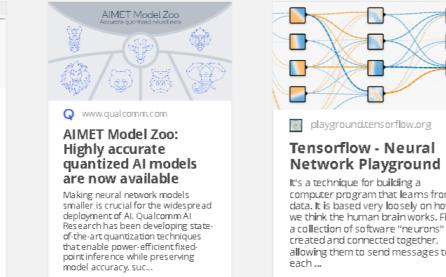


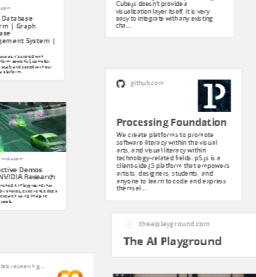




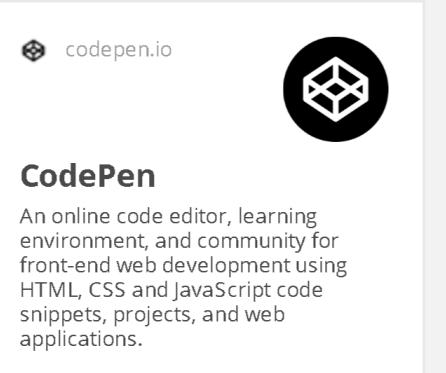


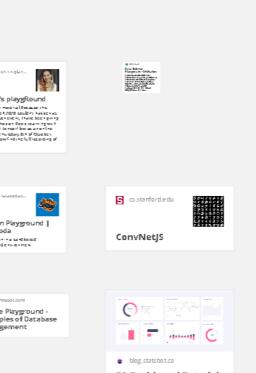




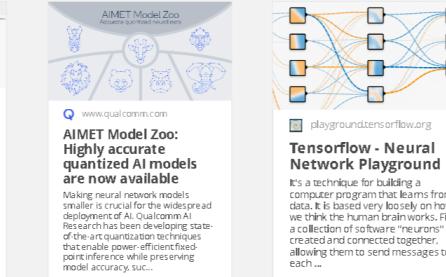


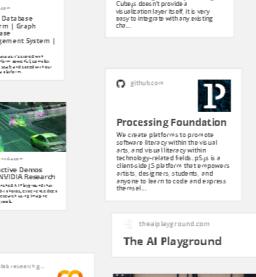




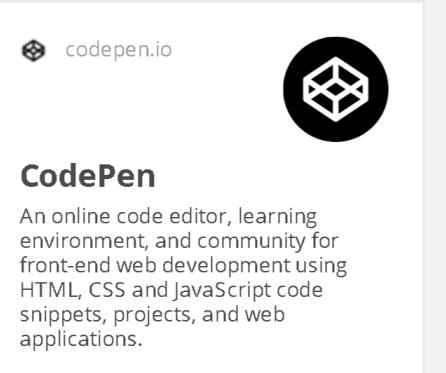


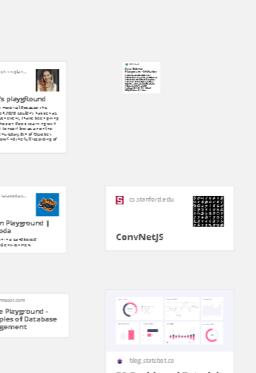




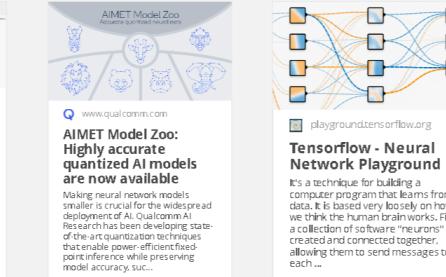


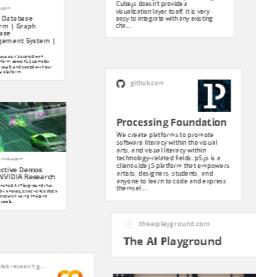




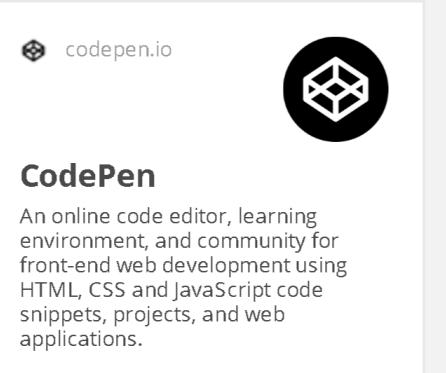


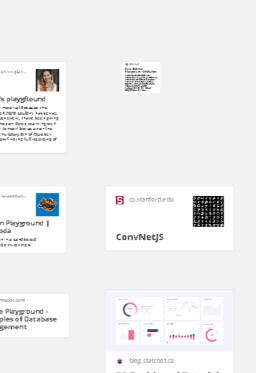




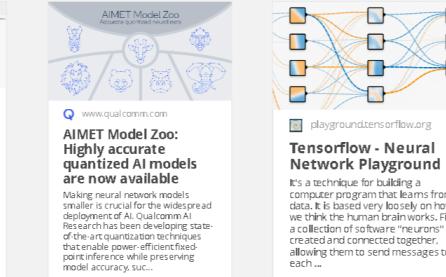


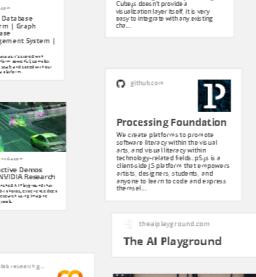




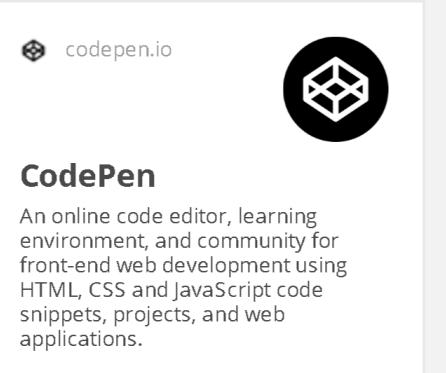


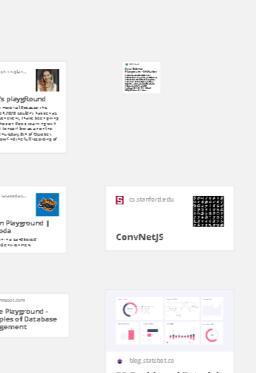




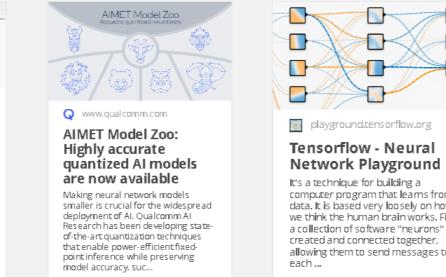


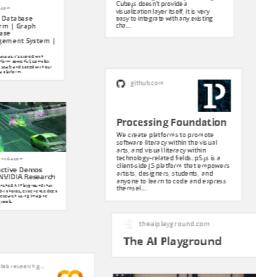




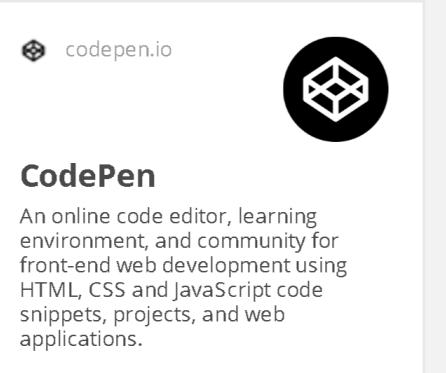


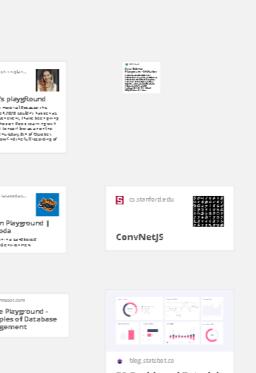




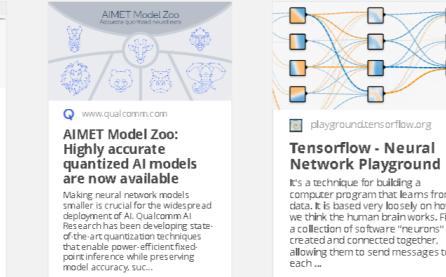


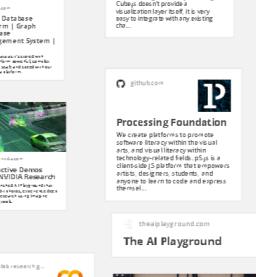


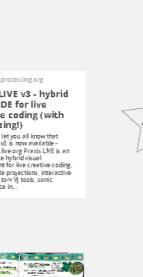


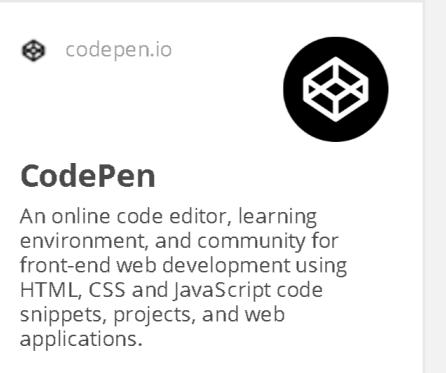


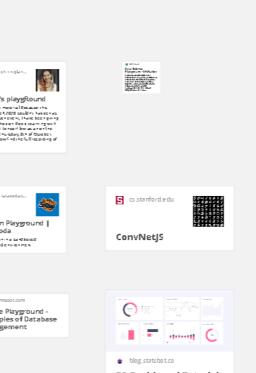




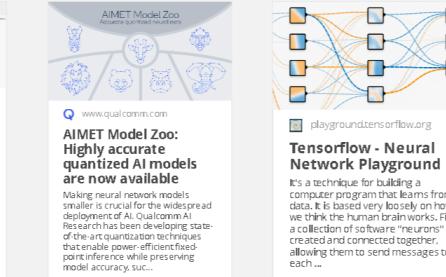


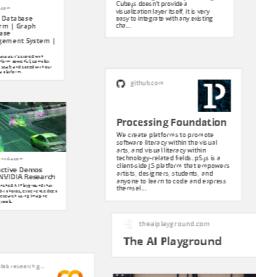




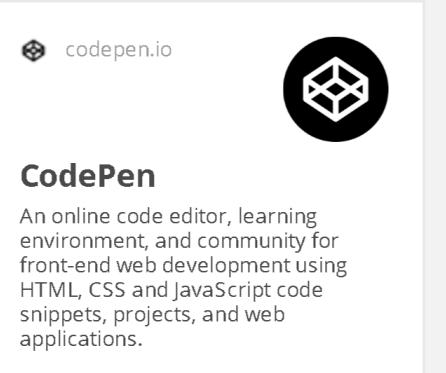


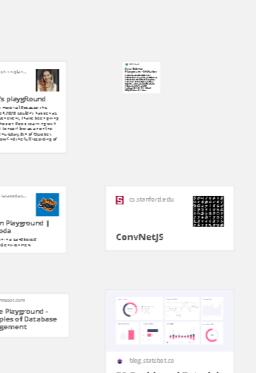




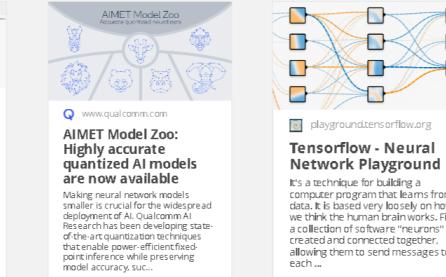


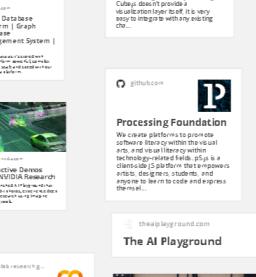


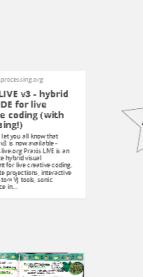


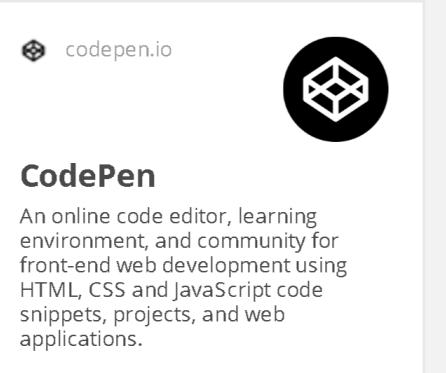


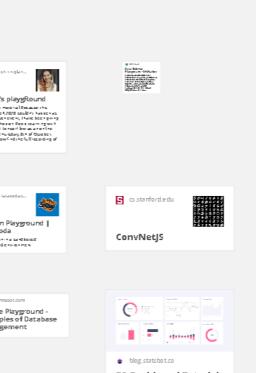




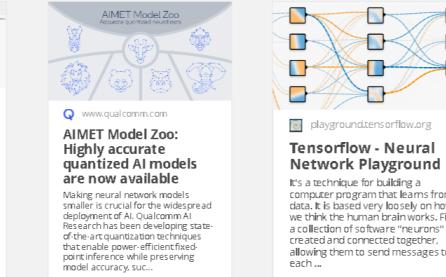


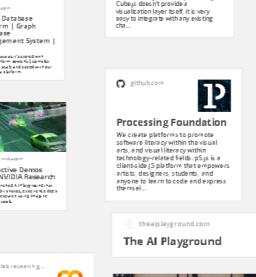




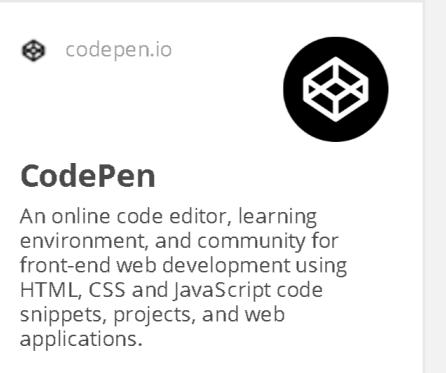


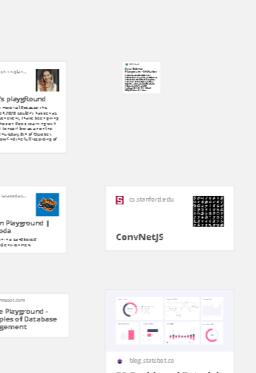




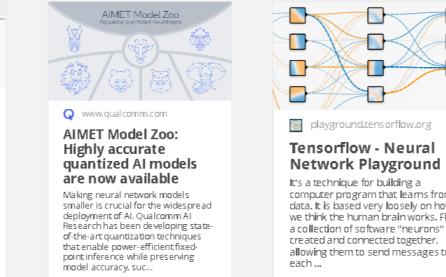


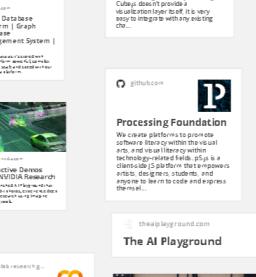




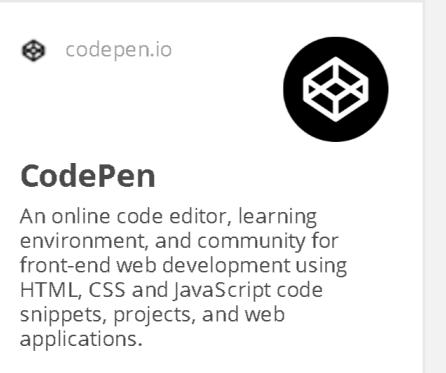


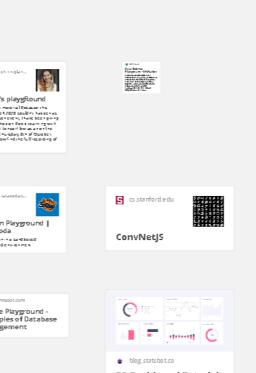




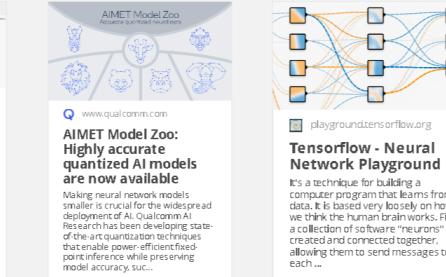


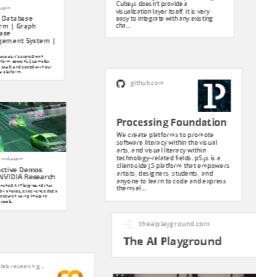




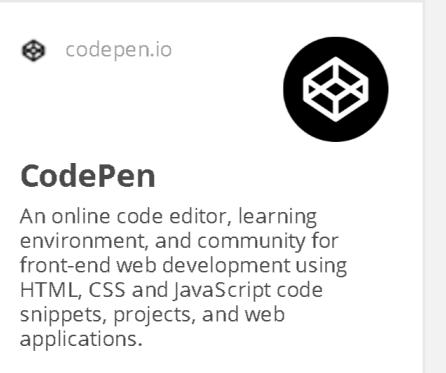


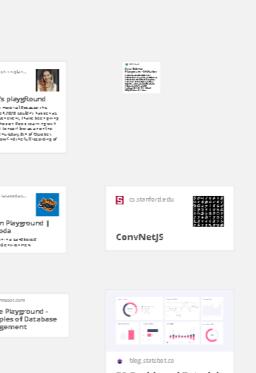




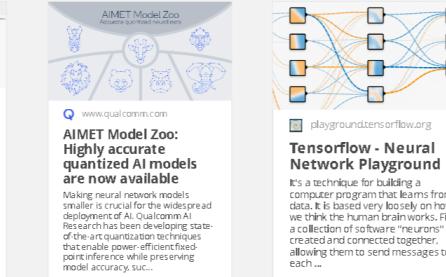


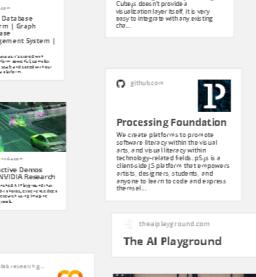




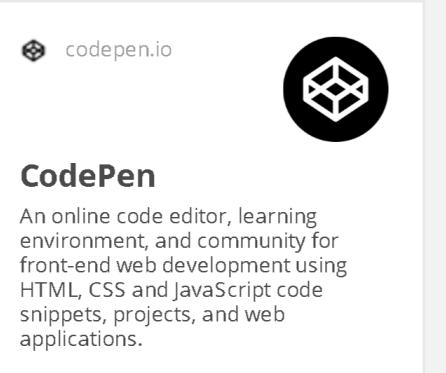


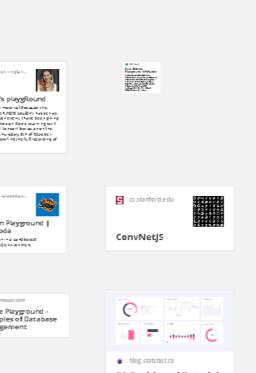




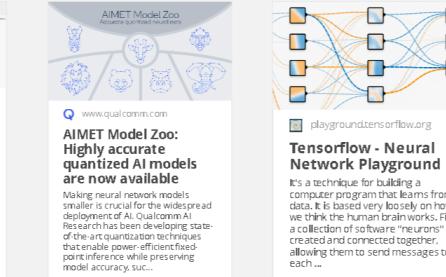


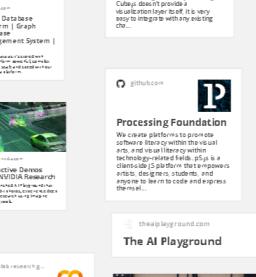




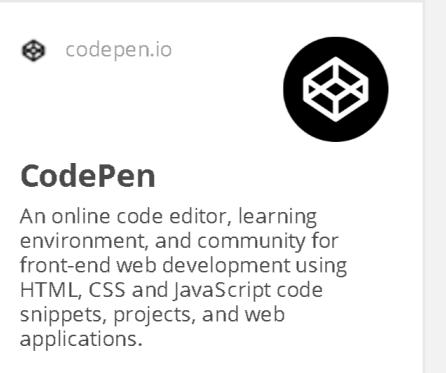


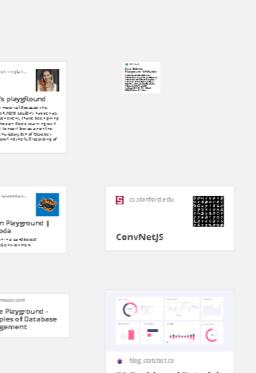




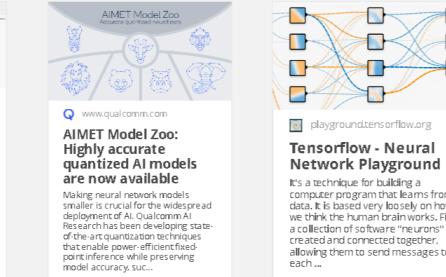


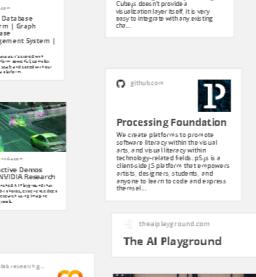




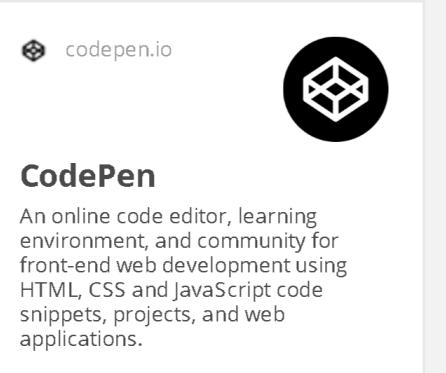


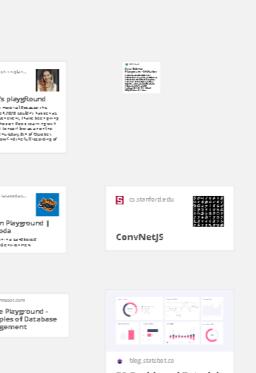




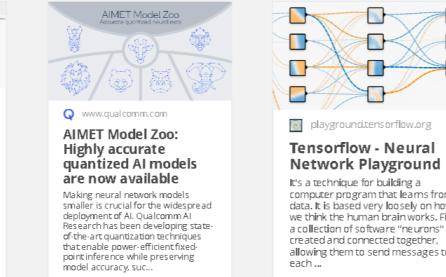


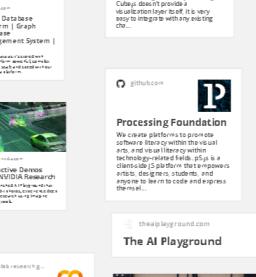


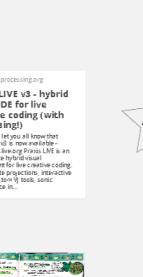


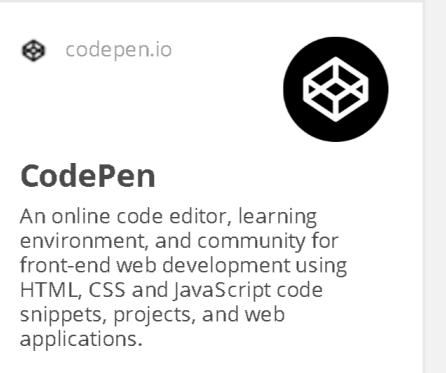


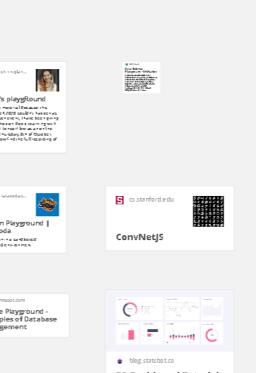




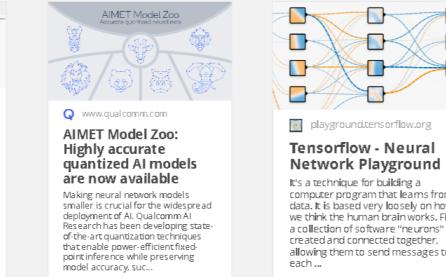


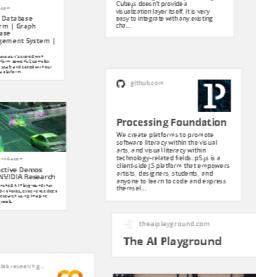




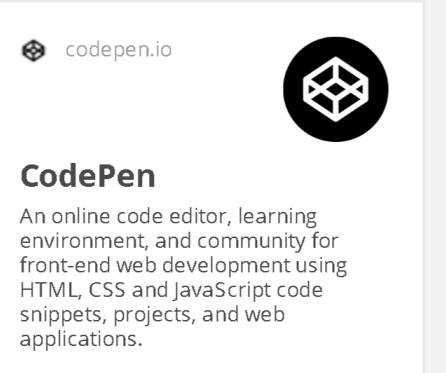


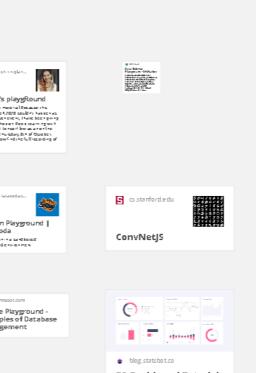




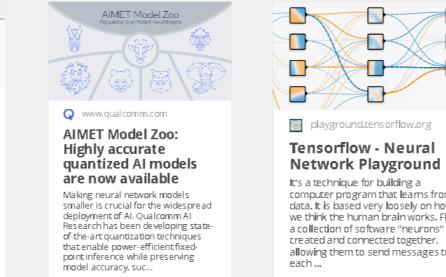


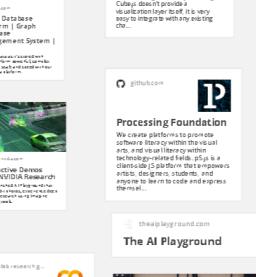




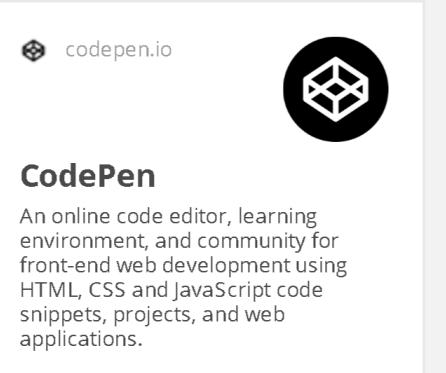


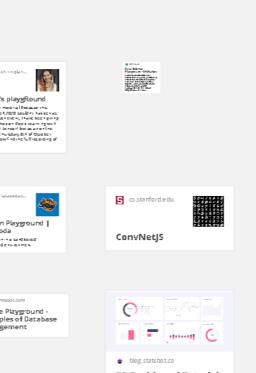




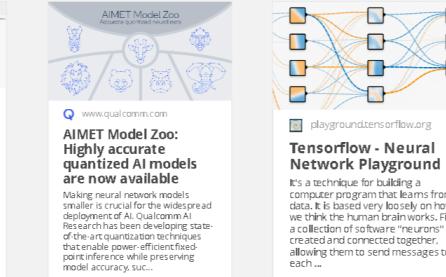


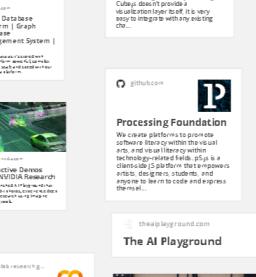




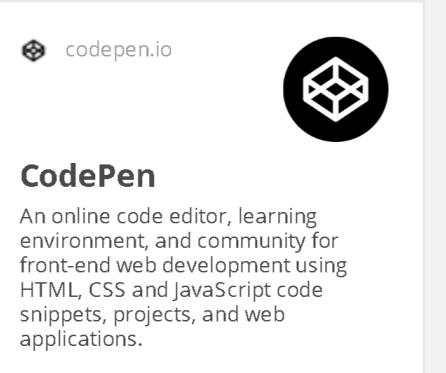


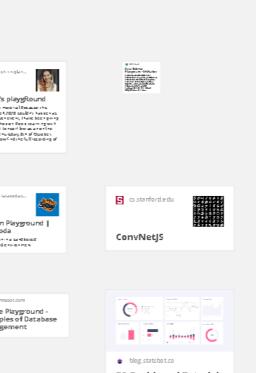




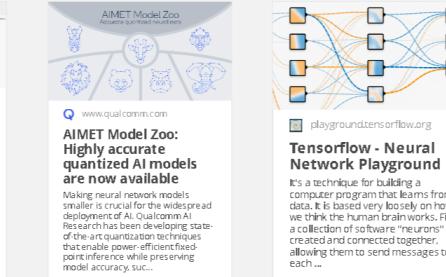


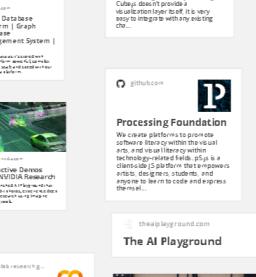




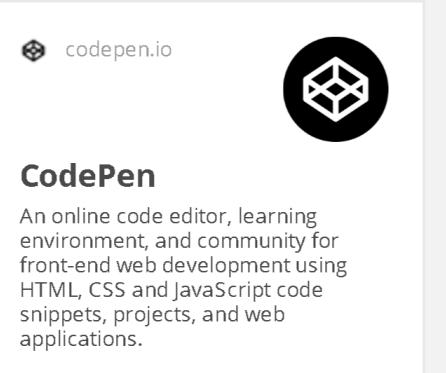


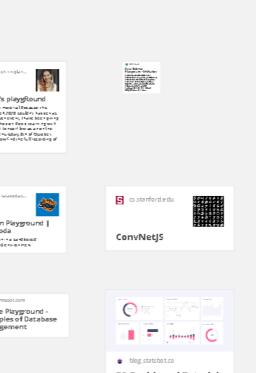




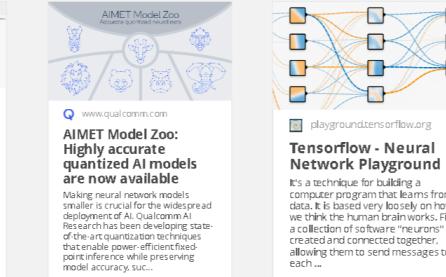


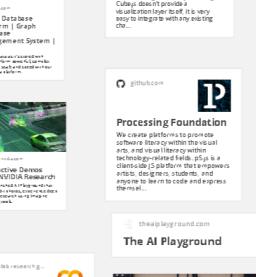




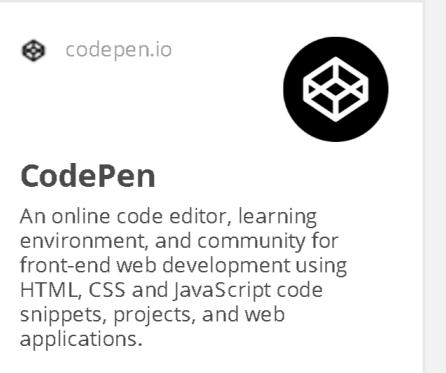


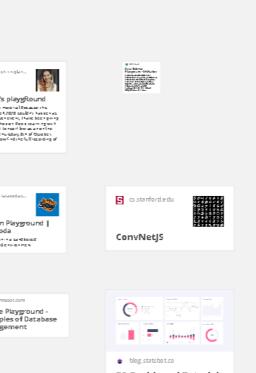














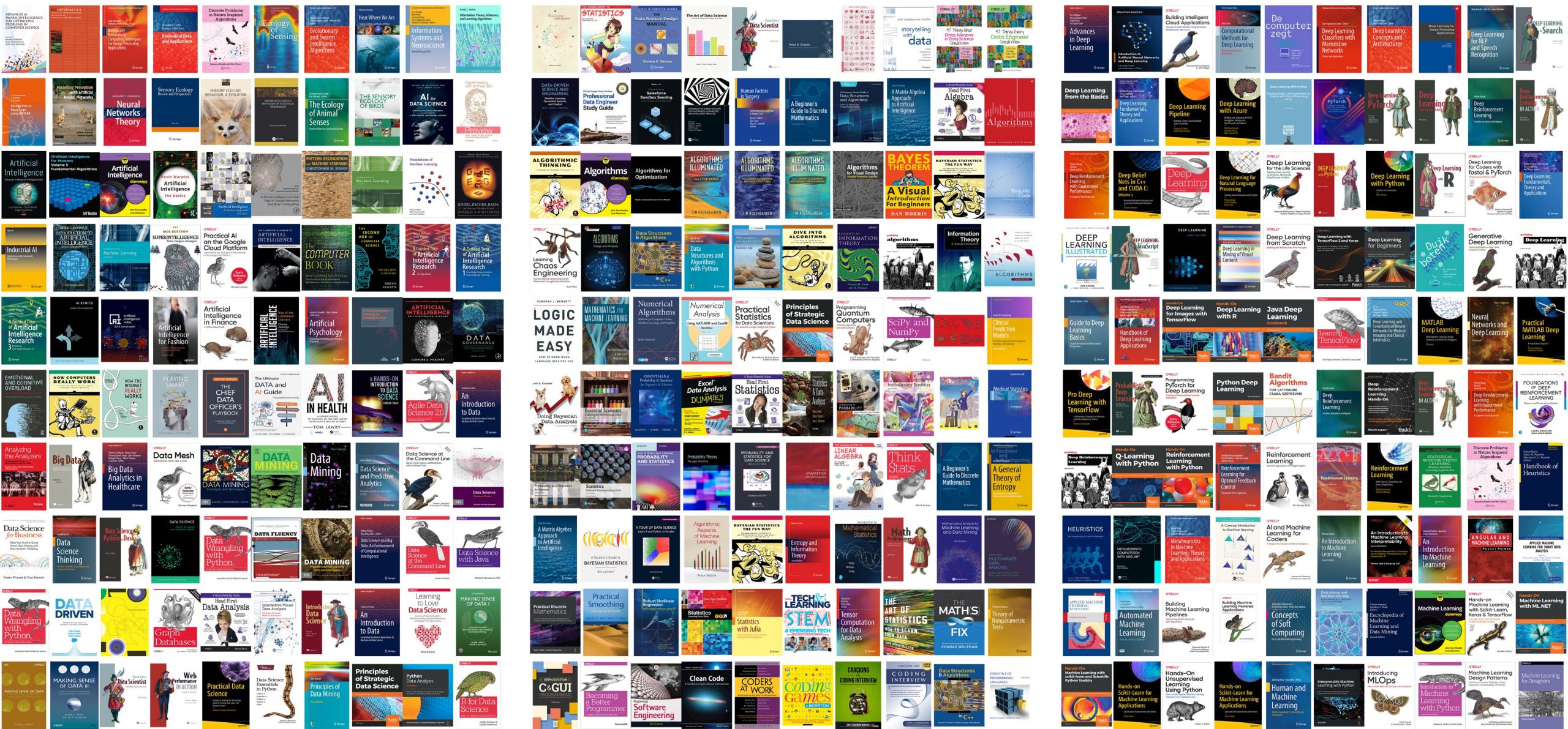
<img alt="Screenshot of the CodePen website

# Learn from prime examples

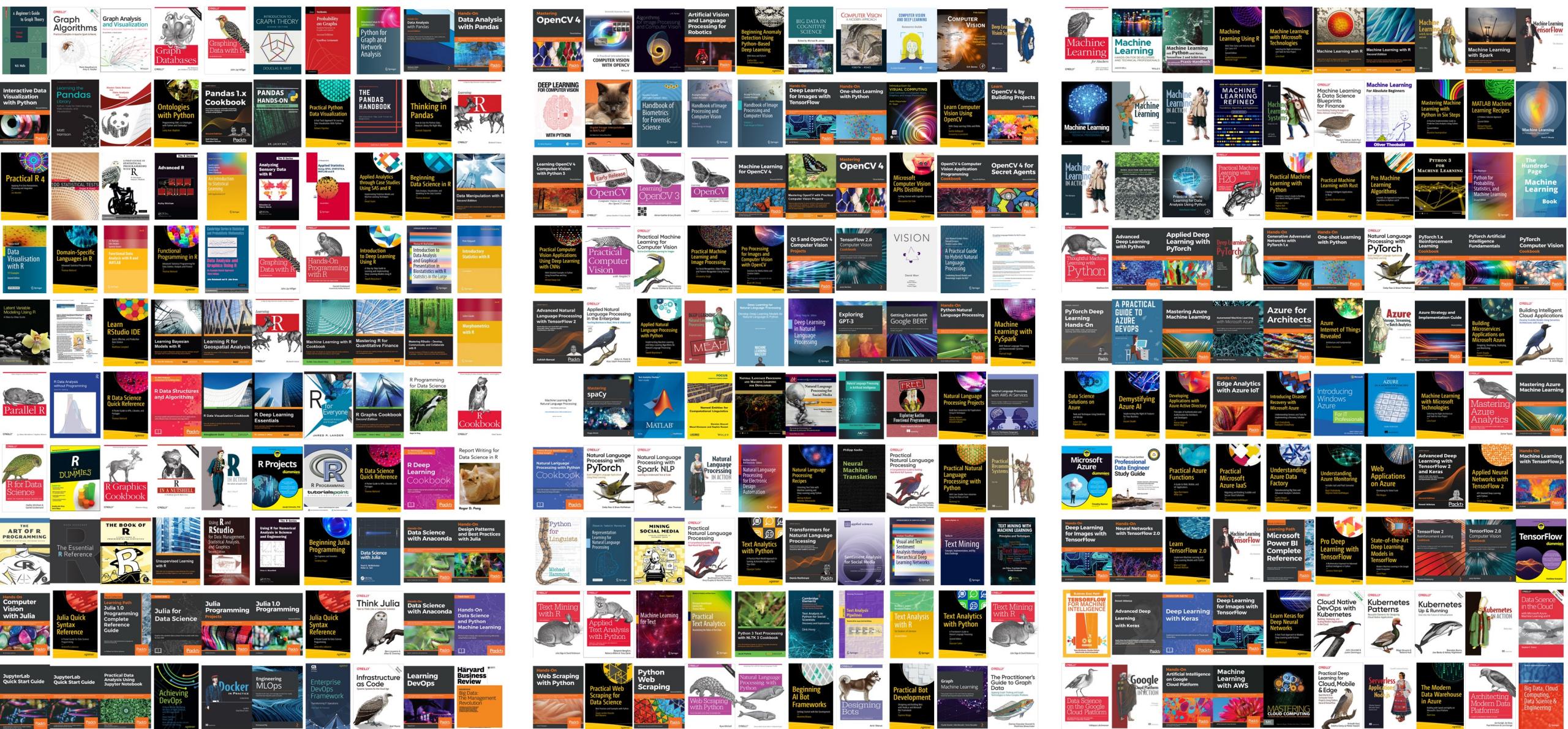
## DATA Science Playgrounds

[https://miro.com/app/board/uXjVOChGvhuE/?invite\\_link\\_id=716932584109](https://miro.com/app/board/uXjVOChGvhuE/?invite_link_id=716932584109)

# {Studied Materials: books}

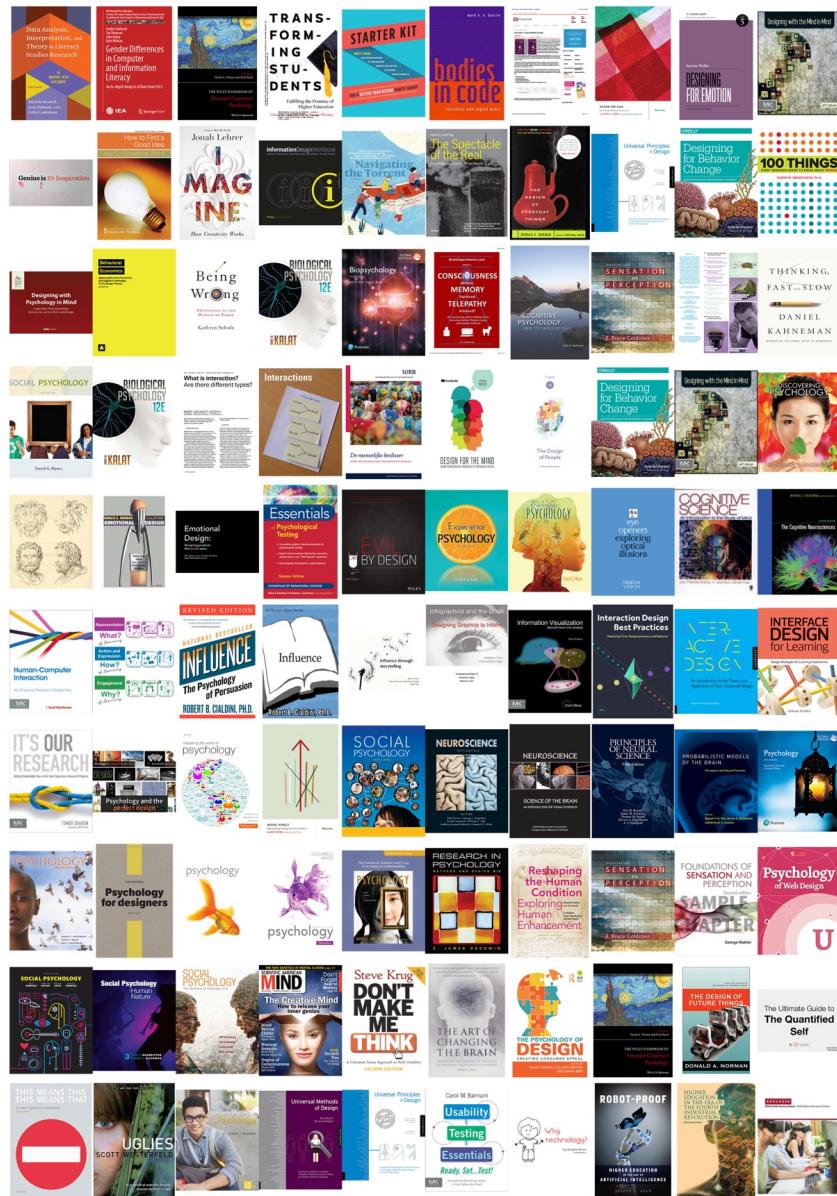


# {Studied Materials: books}



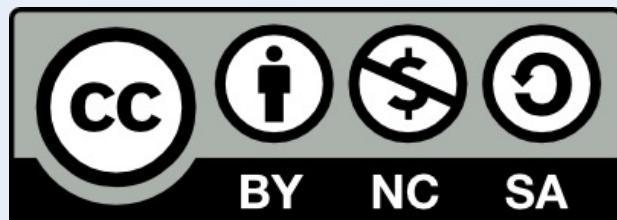
# {Studied Materials: books}

# {Studied Materials: books}

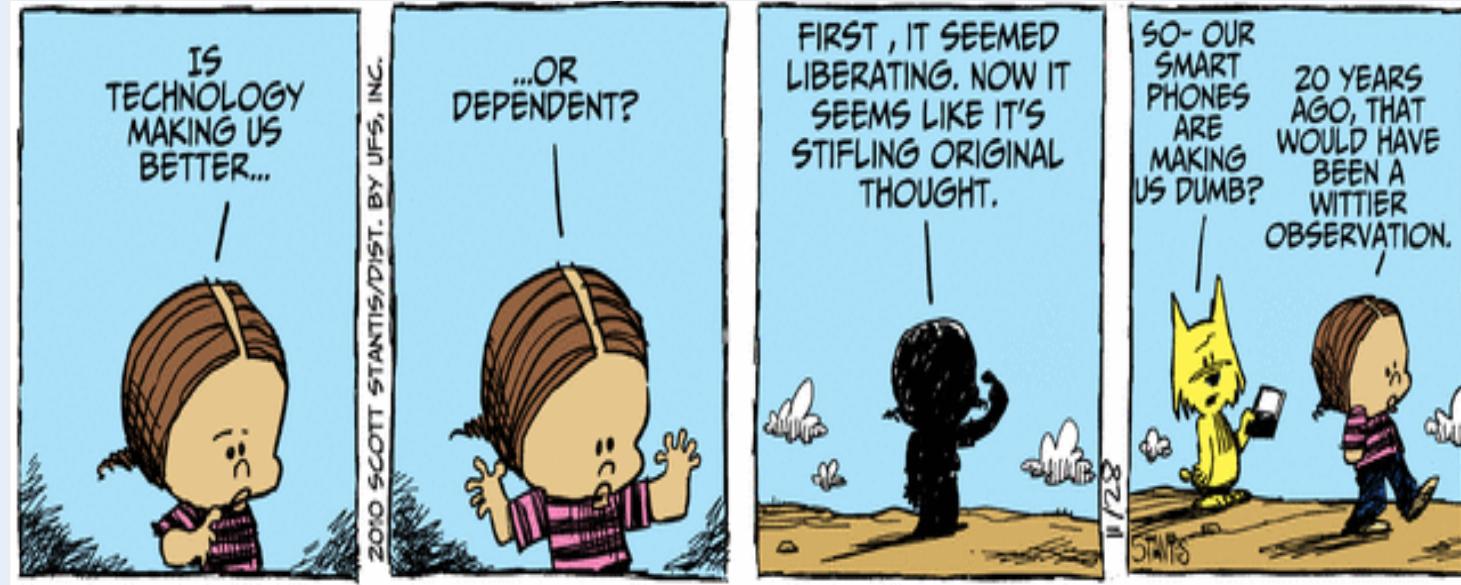


<http://creativecommons.org/licenses/by-nc-sa/3.0/>

These materials are licensed under a Creative Commons Attribution-Share-Alike license. You can change it, transmit it, show it to other people. Just always give credit to RFvdW.



This seminar was developed by:  
**Living-Lab: AiRA,**  
*Hub voor Data & Responsible AI*  
**Rob van der Willigen**  
**JUNE 2022**



Creative Commons License Types		
	Can someone use it commercially?	Can someone create new versions of it?
Attribution	①	
Share Alike	①②	 Yup, AND they must license the new work under a Share Alike license.
No Derivatives	①③	
Non-Commercial	①④	 Yup, AND the new work must be non-commercial, but it can be under any non-commercial license.
Non-Commercial Share Alike	①②③	 Yup, AND they must license the new work under a Non-Commercial Share Alike license.
Non-Commercial No Derivatives	①③④	

SOURCE  
<http://www.masternewmedia.org/how-to-publish-a-book-under-a-creative-commons-license/>