Applied Data Science Project

L4 - Foundation models

Giuseppe Rizzo

Turin, October 3, 2022

On the Opportunities and Risks of Foundation Models

Rishi Bommasani*   Drew A. Hudson   Ehsan Adeli   Russ Altman   Simran Arora
Sydney von Arx   Michael S. Bernstein   Jeannette Bohg   Antoine Bosselut   Emma Brunskill
Erik Brynjolfsson   Shyamal Buch   Dallas Card   Rodrigo Castellon   Niladri Chatterji
Annie Chen   Kathleen Creel   Jared Quincy Davis   Dorottya Demszky   Chris Donahue
Moussa Doumbouya   Esin Durmus   Stefano Ermon   John Etchemendy   Kawin Ethayarajh
Li Fei-Fei   Chelsea Finn   Trevor Gale   Lauren Gillespie   Karan Goel   Noah Goodman
Shelby Grossman   Neel Guha   Tatsunori Hashimoto   Peter Henderson   John Hewitt
Daniel E. Ho   Jenny Hong   Kyle Hsu   Jing Huang   Thomas Icard   Saahil Jain
Dan Jurafsky   Pratyusha Kalluri   Siddharth Karamcheti   Geoff Keeling   Fereshte Khani
Omar Khattab   Pang Wei Koh   Mark Krass   Ranjay Krishna   Rohith Kuditipudi
Ananya Kumar   Faisal Ladhak   Mina Lee   Tony Lee   Jure Leskovec   Isabelle Levent
Xiang Lisa Li   Xuechen Li   Tengyu Ma   Ali Malik   Christopher D. Manning
Suvir Mirchandani   Eric Mitchell   Zanele Munyikwa   Suraj Nair   Avanika Narayan
Deepak Narayanan   Ben Newman   Allen Nie   Juan Carlos Niebles   Hamed Nilforoshan
Julian Nyarko   Giray Ogut   Laurel Orr   Isabel Papadimitriou   Joon Sung Park   Chris Piech
Eva Portelance   Christopher Potts   Aditi Raghunathan   Rob Reich   Hongyu Ren
Frieda Rong   Yusuf Roohani   Camilo Ruiz   Jack Ryan   Christopher Ré   Dorsa Sadigh
Shiori Sagawa   Keshav Santhanam   Andy Shih   Krishnan Srinivasan   Alex Tamkin
Rohan Taori   Armin W. Thomas   Florian Tramèr   Rose E. Wang   William Wang   Bohan Wu
Jiajun Wu   Yuhuai Wu   Sang Michael Xie   Michihiro Yasunaga   Jiaxuan You   Matei Zaharia
Michael Zhang   Tianyi Zhang   Xikun Zhang   Yuhui Zhang   Lucia Zheng   Kaitlyn Zhou
Percy Liang*[1]

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

arXiv:2108.07258v3 [cs.LG] 12 Jul 2022

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical

https://arxiv.org/abs/2108.07258

Reference paper

On the Opportunities and Risks of Foundation Models

Rishi Bommasani*   Drew A. Hudson   Ehsan Adeli   Russ Altman   Simran Arora
Sydney von Arx   Michael S. Bernstein   Jeannette Bohg   Antoine Bosselut   Emma Brunskill
Erik Brynjolfsson   Shyamal Buch   Dallas Card   Rodrigo Castellon   Niladri Chatterji
Annie Chen   Kathleen Creel   Jared Quincy Davis   Dorottya Demszky   Chris Donahue
Moussa Doumbouya   Esin Durmus   Stefano Ermon   John Etchemendy   Kawin Ethayarajh
Li Fei-Fei   Chelsea Finn   Trevor Gale   Lauren Gillespie   Karan Goel   Noah Goodman
Shelby Grossman   Neel Guha   Tatsunori Hashimoto   Peter Henderson   John Hewitt
Daniel E. Ho   Jenny Hong   Kyle Hsu   Jing Huang   Thomas Icard   Saahil Jain
Dan Jurafsky   Pratyusha Kalluri   Siddharth Karamcheti   Geoff Keeling   Fereshte Khani
Omar Khattab   Pang Wei Koh   Mark Krass   Ranjay Krishna   Rohith Kuditipudi
Ananya Kumar   Faisal Ladhak   Mina Lee   Tony Lee   Jure Leskovec   Isabelle Levent
Xiang Lisa Li   Xuechen Li   Tengyu Ma   Ali Malik   Christopher D. Manning
Suvir Mirchandani   Eric Mitchell   Zanele Munyikwa   Suraj Nair   Avanika Narayan
Deepak Narayanan   Ben Newman   Allen Nie   Juan Carlos Niebles   Hamed Nilforoshan
Julian Nyarko   Giray Ogut   Laurel Orr   Isabel Papadimitriou   Joon Sung Park   Chris Piech

[cs.LG]   12 Jul 2022

arXiv

> "The beginning of a paradigm shift: foundation models have only just begun to transform the way AI systems are built and deployed in the world"

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical

https://arxiv.org/abs/2108.07258

Reference paper

On the Opportunities and Risks of Foundation Models

Rishi Bommasani*   Drew A. Hudson   Ehsan Adeli   Russ Altman   Simran Arora
Sydney von Arx   Michael S. Bernstein   Jeannette Bohg   Antoine Bosselut   Emma Brunskill
Erik Brynjolfsson   Shyamal Buch   Dallas Card   Rodrigo Castellon   Niladri Chatterji
Annie Chen   Kathleen Creel   Jared Quincy Davis   Dorottya Demszky   Chris Donahue
Moussa Doumbouya   Esin Durmus   Stefano Ermon   John Etchemendy   Kawin Ethayarajh
Li Fei-Fei   Chelsea Finn   Trevor Gale   Lauren Gillespie   Karan Goel   Noah Goodman
Shelby Grossman   Neel Guha   Tatsunori Hashimoto   Peter Henderson   John Hewitt
Daniel E. Ho   Jenny Hong   Kyle Hsu   Jing Huang   Thomas Icard   Saahil Jain
Dan Jurafsky   Pratyusha Kalluri   Siddharth Karamcheti   Geoff Keeling   Fereshte Khani
Omar Khattab   Pang Wei Koh   Mark Krass   Ranjay Krishna   Rohith Kuditipudi
Ananya Kumar   Faisal Ladhak   Mina Lee   Tony Lee   Jure Leskovec   Isabelle Levent
Xiang Lisa Li   Xuechen Li   Tengyu Ma   Ali Malik   Christopher D. Manning
Suvir Mirchandani   Eric Mitchell   Zanele Munyikwa   Suraj Nair   Avanika Narayan
Deepak Narayanan   Ben Newman   Allen Nie   Juan Carlos Niebles   Hamed Nilforoshan
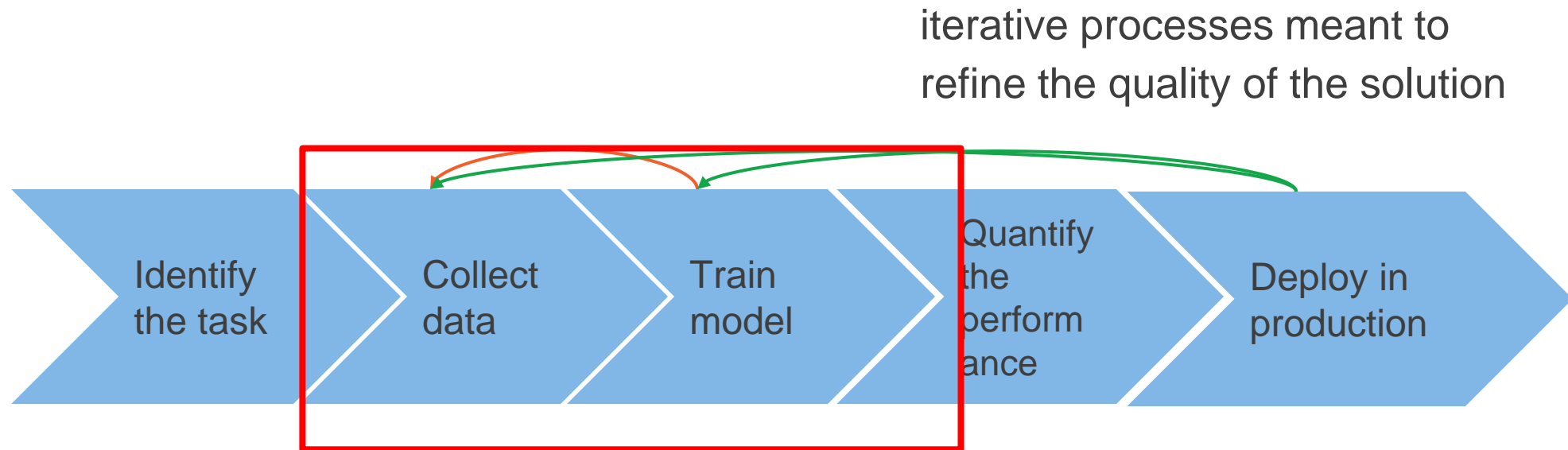Julian Nyarko   Giray Ogut   Laurel Orr   Isabel Papadimitriou   Joon Sung Park   Chris Piech

[cs.LG] 12 Jul 2022

"A foundation model is any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks … examples include BERT, GPT-3, and CLIP."
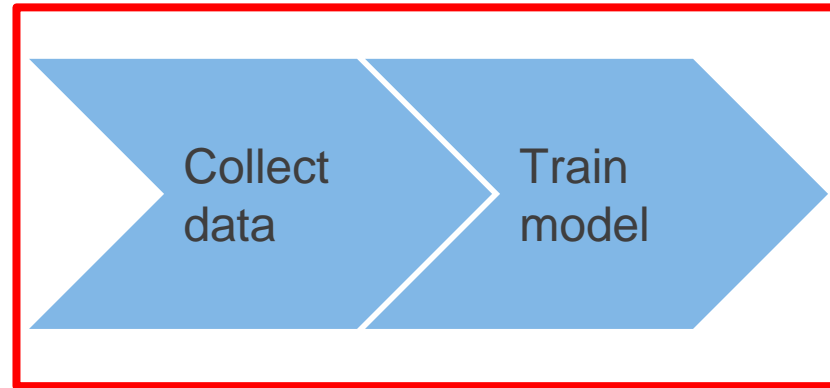
AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical

arXiv

https://arxiv.org/abs/2108.07258

(Today) artificial intelligence

iterative processes meant to refine the quality of the solution

Identify the task

Collect data

Train model

Quantify the performance
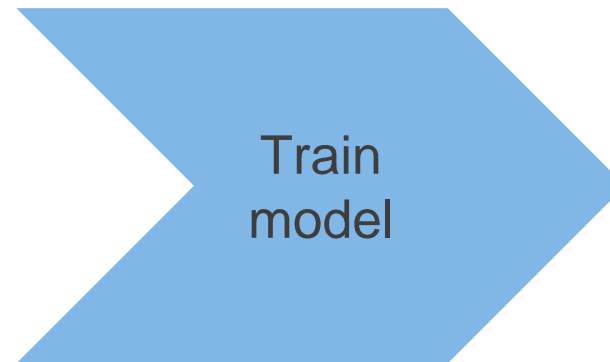
Deploy in production

Model is the output of 2 sequential activities

model is obtained after the data is collected, and the algorithm is repeated iteratively

Should intelligence start from scratch always?

Train
model

Re-use existing already trained models and adapt or tune to address specific scenarios

Foundation models

Re-use & Adapt

Foundation models are defined in the scope of the machine learning field

It is a similar concept than <u>pre-trained</u>, <u>self-supervision</u>, <u>fine-tuning</u>, or <u>transfer learning</u>

The term foundation inherits both the intrinsic value of a basic component plus the uncomplete essence of the component, yet unfinished

Definition

A foundation model is a large artificial intelligence model trained on a vast quantity of (unlabelled) data at scale (usually by self-supervised learning) resulting in a model that can be adapted to a wide range of downstream tasks
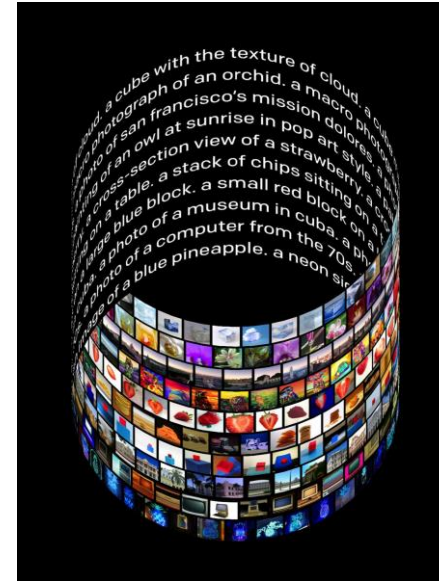
Foundation models
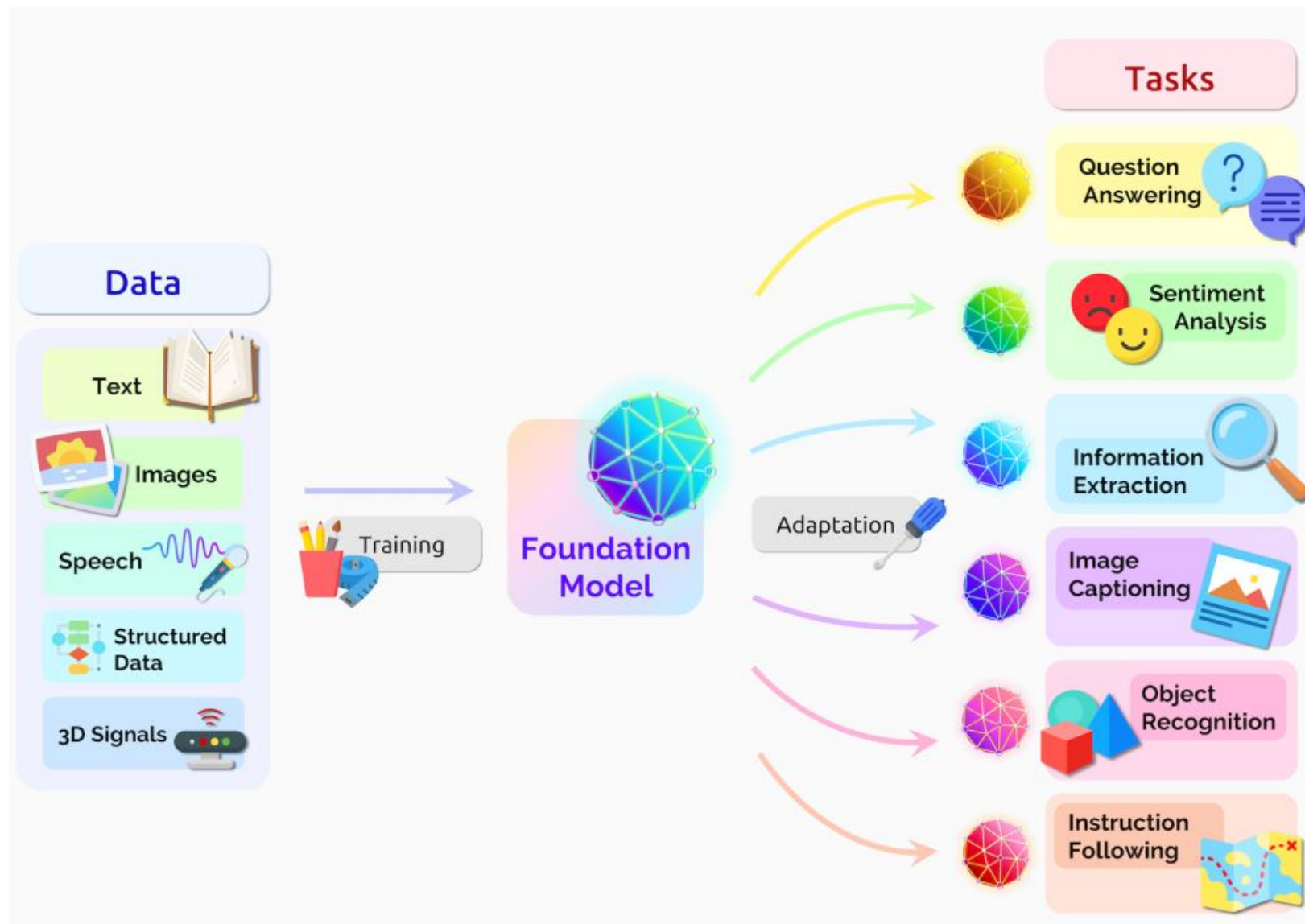
Foundation models

Foundation models are scientifically interesting due to their impressive performance and capabilities

linksfoundation.com
COPYRIGHT ©2021 LINKS

Examples

BERT, 100M+ parameters
Language model

DALL-E, 10M+ parameters
Text descriptions -> images

Who generates foundation models today?

It is a business of a handful of very large companies with very large resource capabilities such as Google, Facebook, Microsoft, Huawei

Two startups are part of this business namely OpenAI, AI21 Labs with significant resource facilities

What about others universities, research centers, other organizations such as companies locally?

They simply cannot keep up with what these players are generating, because they do not possess the same resources namely computing power, data at disposal

Availabilities

Are foundation models available?

Often yes and usually there are available both source codes, models, reports

Some examples:

- BERT https://github.com/google-research/bert
- DALL-E https://github.com/openai/DALL-E
- CLIP https://github.com/openai/CLIP

GPT-3

It is only available the report

They are not released both the source code and models

It can be tested and utilized via the API https://beta.openai.com

Standing on the shoulders of giants

It has been a common practise since the advent of humanity mostly

It is considered a default in research and adopted by most of the researchers worldwide

It is a value given back to practitioners and society for the predominant position they acquired in these years for this business and the digital business as whole

Social impact and economic consequences

The availability of (some) foundation models makes critical to study because they are quickly integrated into real world AI systems

Short analysis of their implications on society:

- do they discriminate?

- do they favour certain user groups?

On the other side, a significant economic consequence: it is now possible to wrap it up a service that implements a task in a level of precision and at a scale that before was not possible
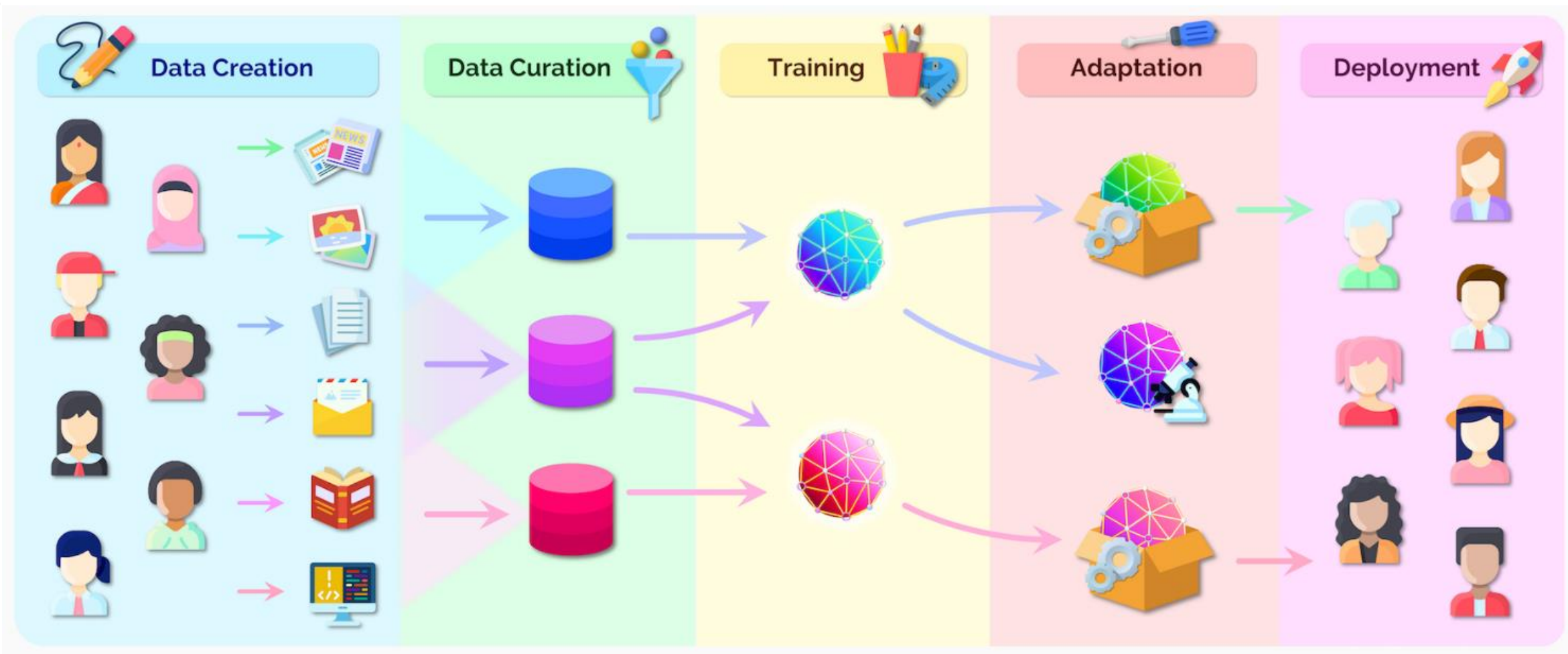
linksfoundation.com

Opportunities

- Having at disposal Google-level performance in text analysis, generation, image rendering with limited resources

- Reducing the impact of the generation of machine intelligence to environment and people

- Learning from examples and fostering a culture of AI

- Fostering a culture of these exact tools, shedding light on key aspects but also risks
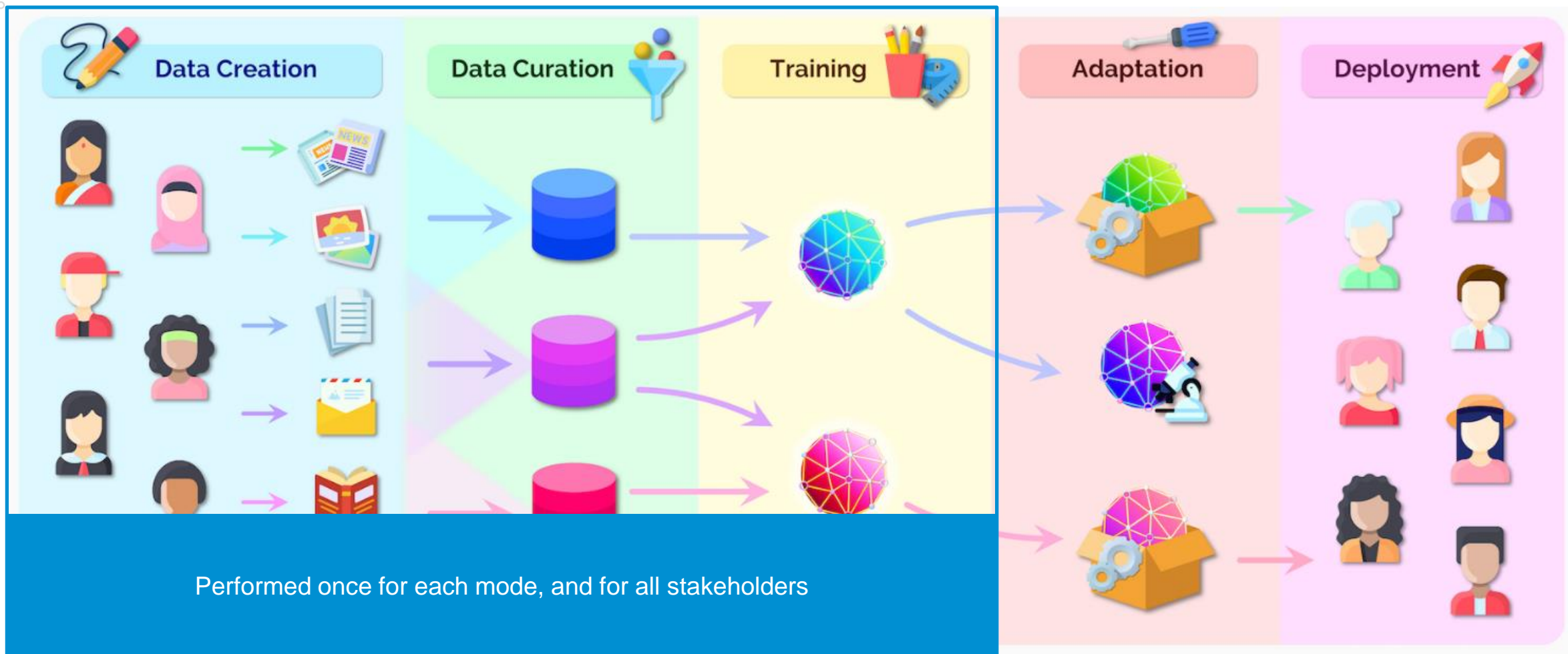
Risks

- Inequity and fairness
    - furthering the unjust treatment of people who have been historically discriminated against
    - lack of diversity in the training examples
- Misuse
    - utilizing the models to harm people (not the original intent of their design)
- Environment
    - increasing pollution in their making
- Legality
    - who is responsible for a wrong action or decision?
- Economics
    - benefits spread not just in the hands of the giants
- Ethics of scale
    - homogenizing decisions, lowering the diversity in the making and the acceptance

The new value chain

Data Creation → Data Curation → Training → Adaptation → Deployment

The new value chain

Data Creation | Data Curation | Training | Adaptation | Deployment

Performed once for each mode, and for all stakeholders

Foundation models ecosystem

Data Creation

It has a dual impact because any foundation model needs good data to be bootstrapped, and its adaptation needs good data to be tailored to the specific task

For instance, if we aim to identify sentiment of text, we may use a foundation that is rooted in a collection of newswire contents, but then we need examples of reviews
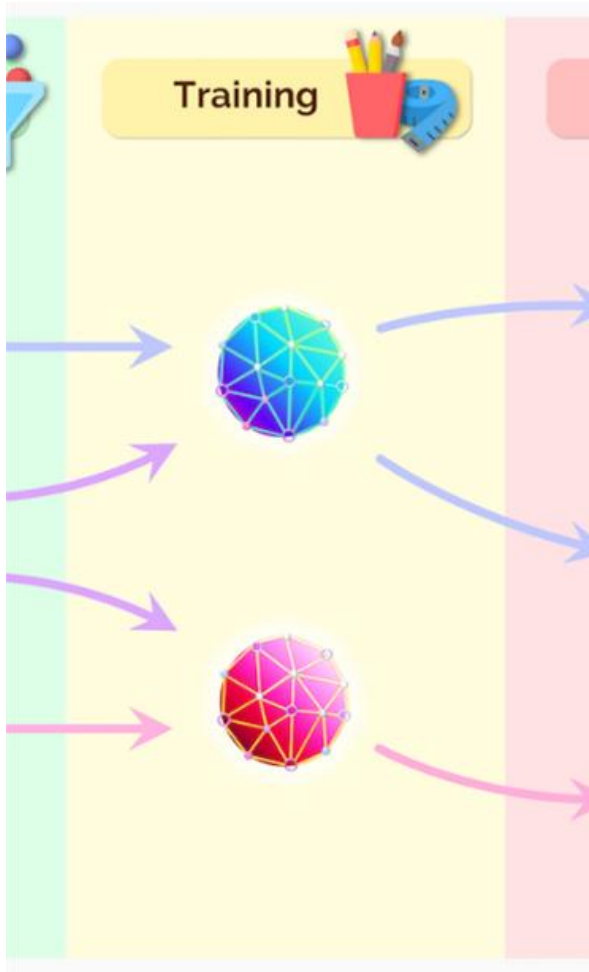
Foundation models ecosystem
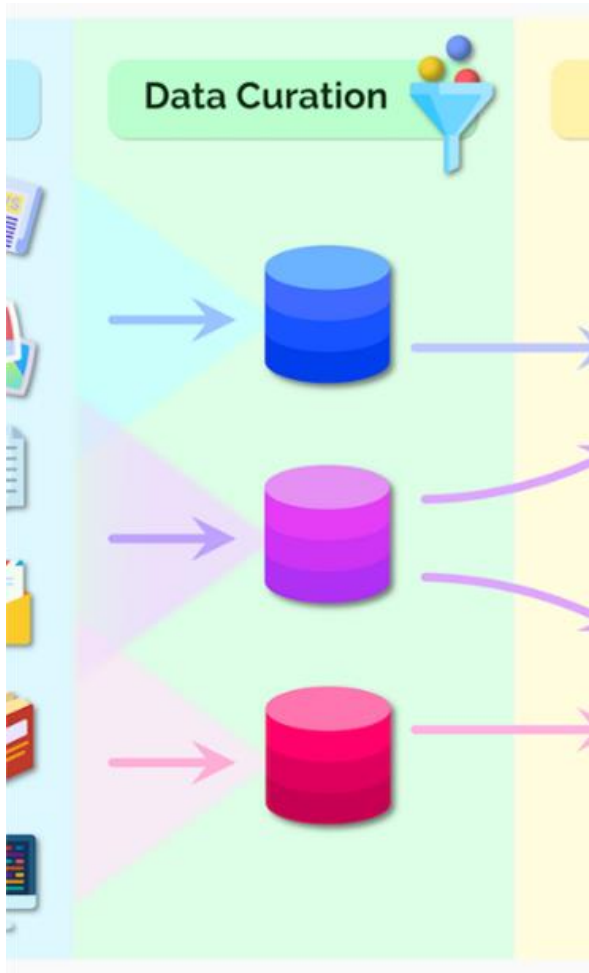
Dual impact given the need for data curation both for creating a consistent and complete dataset

If our target is to classify images according to emotions, we will first need a dataset of images with labels of objects and then a set of images rated by emotions
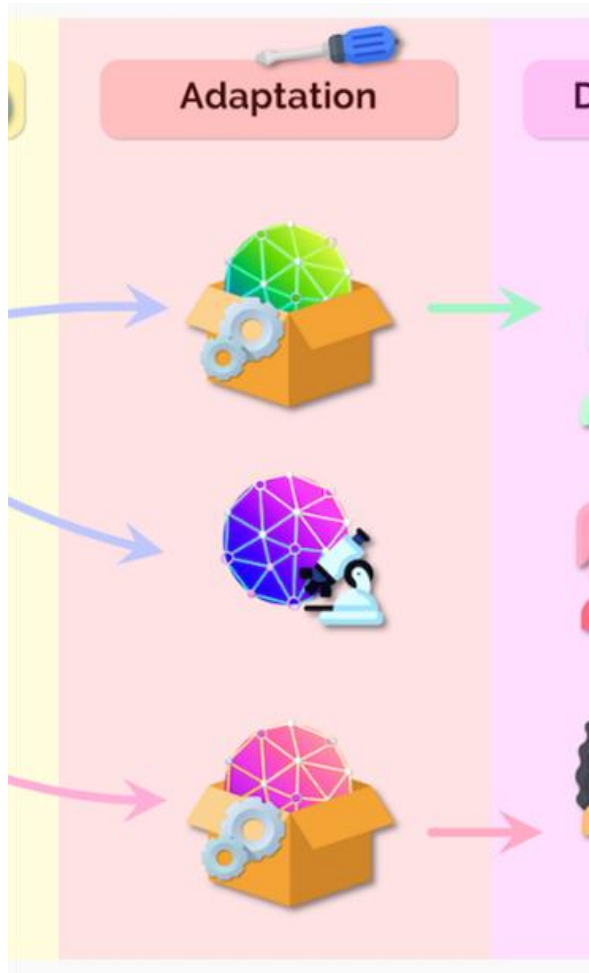
Foundation models ecosystem

Training is a step required to generate the foundation model

For instance, starting from a set of newswire content we will package a model by instructing the software to repeat the task by indexing all examples present in the dataset

Foundation models ecosystem

It is a step required to generate a task specific (narrow) intelligence by adapting the (broader) intelligence represented by the foundation model

For instance, it is about leveraging the collected dataset in a training procedure meant to specialize the foundation model. In practise: it adds up some additional components (for instance neural network layers) to rework the input of the foundation model to address the targeted task

Foundation models ecosystem
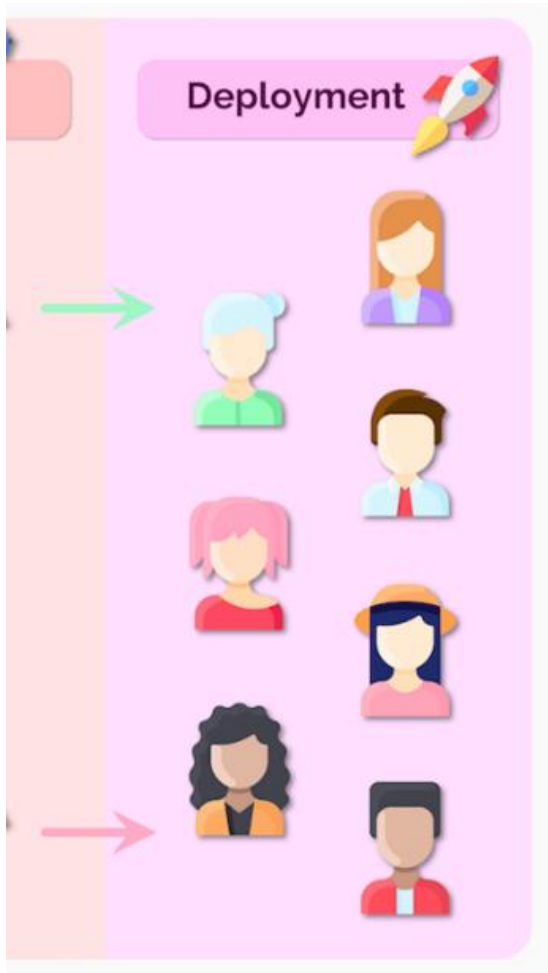
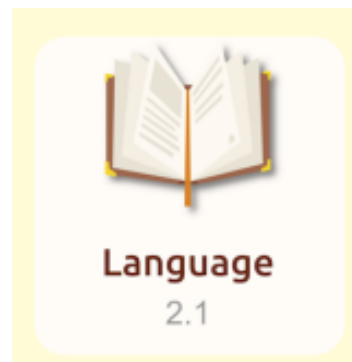It is about reaching the users of the models by offering them access to the intelligence developed

For instance, users of a web application that helps users to compute the number of positive vs negative reviews of their products

Capabilities

What foundation models can offer?

Language 2.1 · Vision 2.2 · Robotics 2.3 · Reasoning 2.4 · Interaction 2.5 · Philosophy 2.6

Capabilities

Language
2.1

Predict the next word

Translate text

Identify from text key relevant information

Recognize sentiment

...

Capabilities

Vision

2.2

Recognize objects

Recognize face expressions

Recognize emotions

...

Capabilities

Routing in an closed unknown environment

Physical understanding

...

Capabilities

First stub content generation (news, source code)

Multimodal interaction (voice, vision)

...

Capabilities

Philosophy
2.6

Truly and deep understanding of the environment (such as meanings of objects and their intertwinings)

...

Thank you for your attention.

Questions?

CONTACTS

Giuseppe Rizzo

Program Manager (LINKS Foundation) and
Adjunct Professor (Politecnico di Torino)

giuseppe.rizzo@polito.it