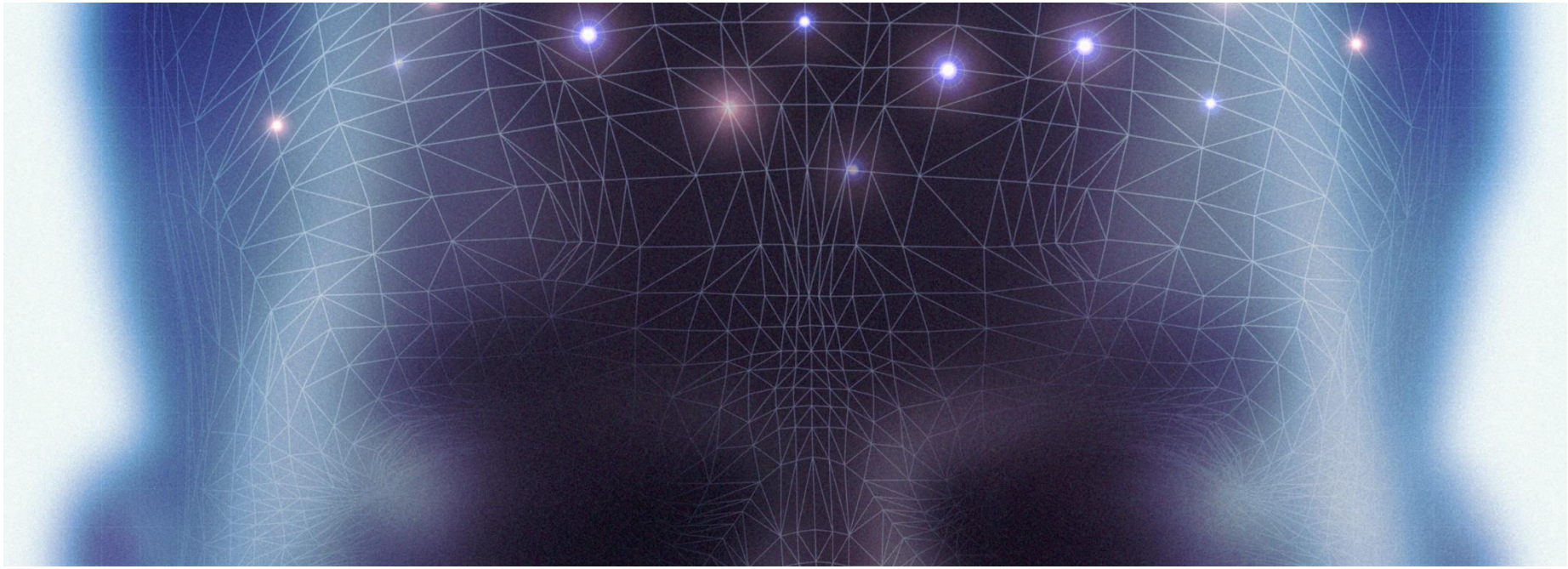
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für Verteidigung,
Bevölkerungsschutz und Sport VBS
Bundesamt für Rüstung - armasuisse ar



Automated Identification of Emerging Technologies: Open Data Approach

Ljiljana Dolamic, **Julian Jang-Jaccard**, Alain Mermoud, and Vincent Lenders
Cyber-Defence Campus, armasuisse Science and Technology, Switzerland

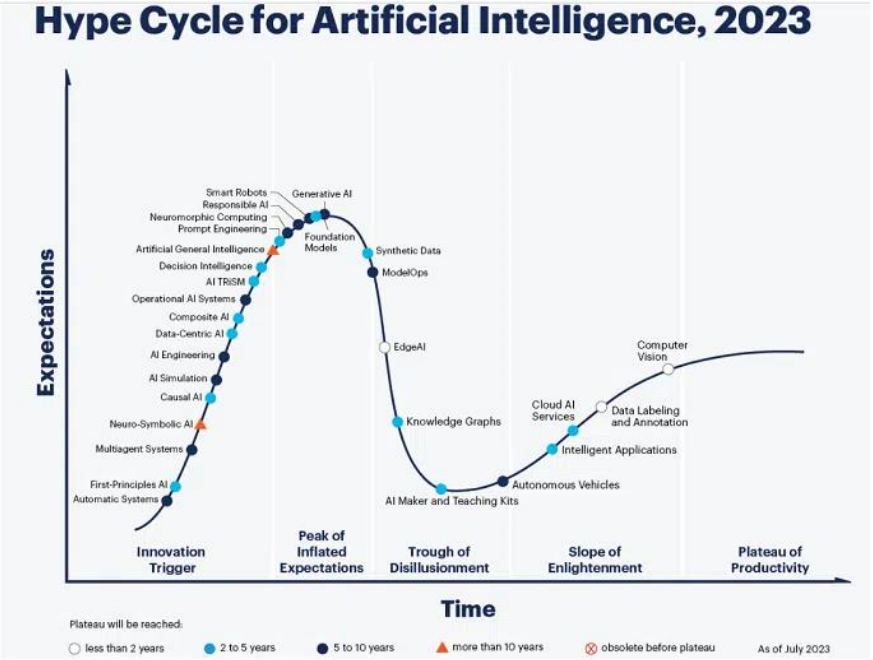
23 April 2024

Setting the Scene

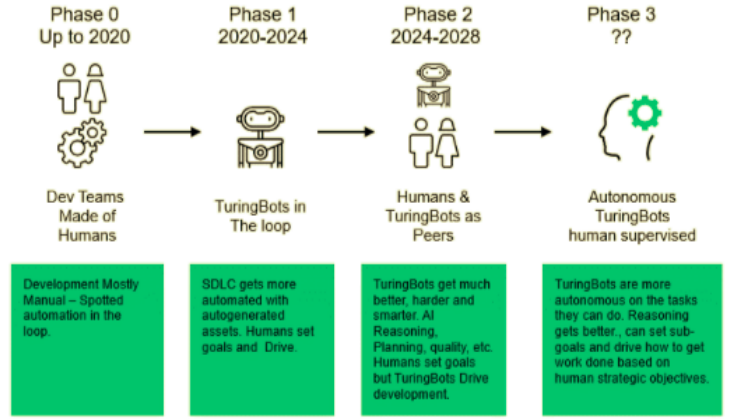
- Understanding emerging technologies is crucial for various entities, e.g, industry, academia, and government agencies.
- No single standard agreement on what constitutes the term?
- Makes it challenging to develop a **scientifically sound** methodology to **identify emerging technologies**.

Existing Approaches

- Relies on qualitative methods, e.g., expert opinion, survey-based.
- For quantitative methods, open datasets and S-curve models but mostly on specific predetermined sets of technologies (e.g, IoT, chemistry)
- Market research firms (e.g., Gartner, Forrester) produce annual reports, yet the methodology remains unclear.



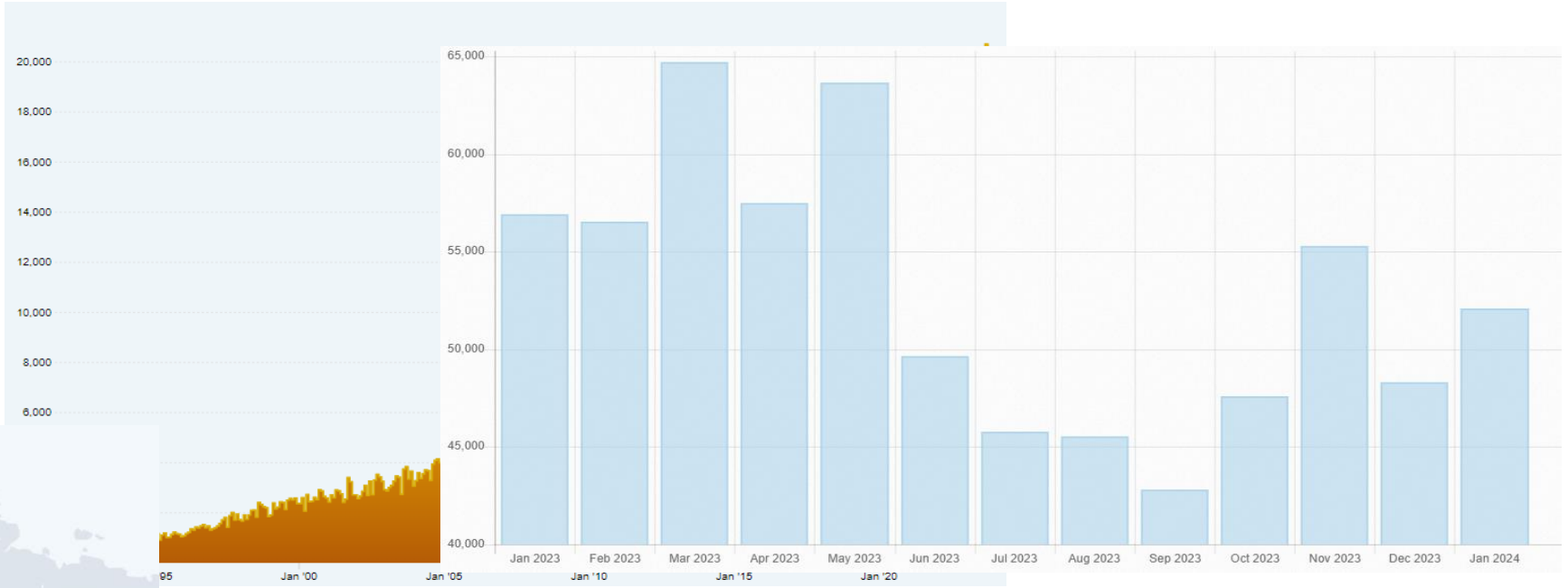
Source: Gartner https://www.gartner.com/

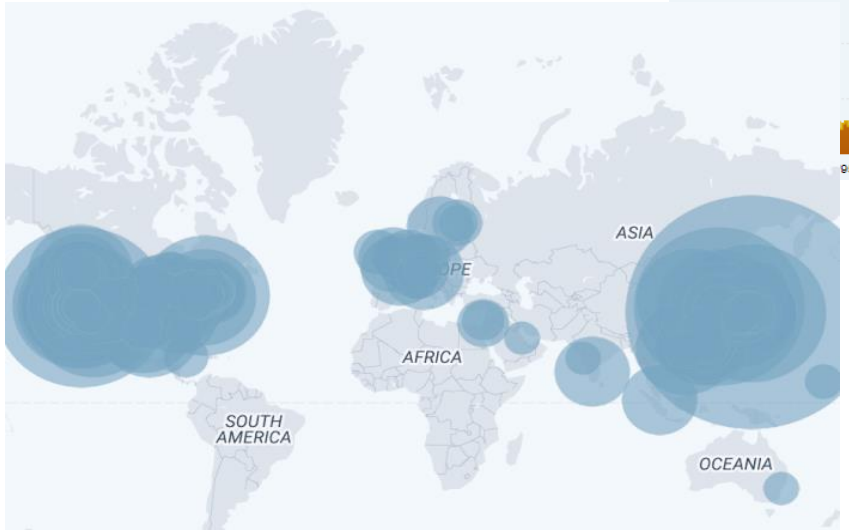


Source: Forrester https://www.forrester.com/

Quantatitve Open Data Approach

Interest of Research Community:
Publications from arXiv - focuses on a subset of approx. 1.4 million arXiv publications.

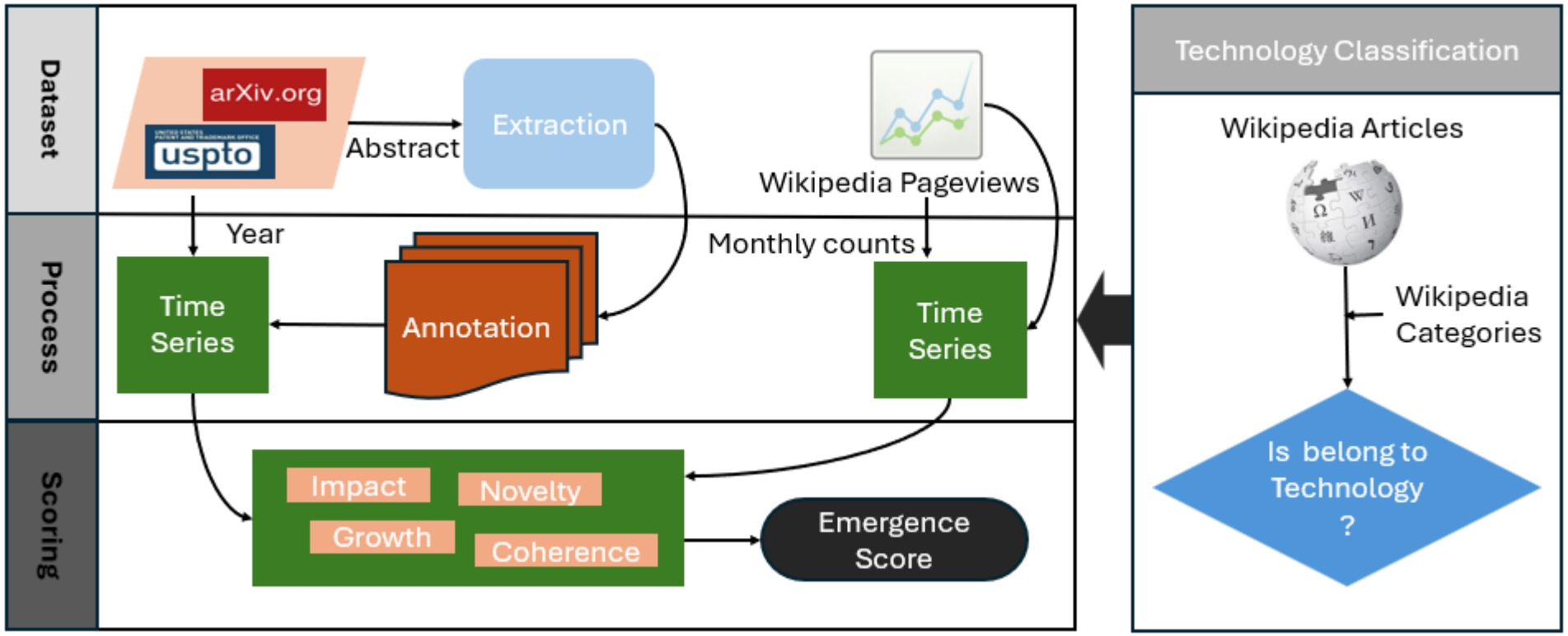




Interest of Innovation Community:
Patents from PatentsView - utilize a subset of around 6.6 million patent records for our study.

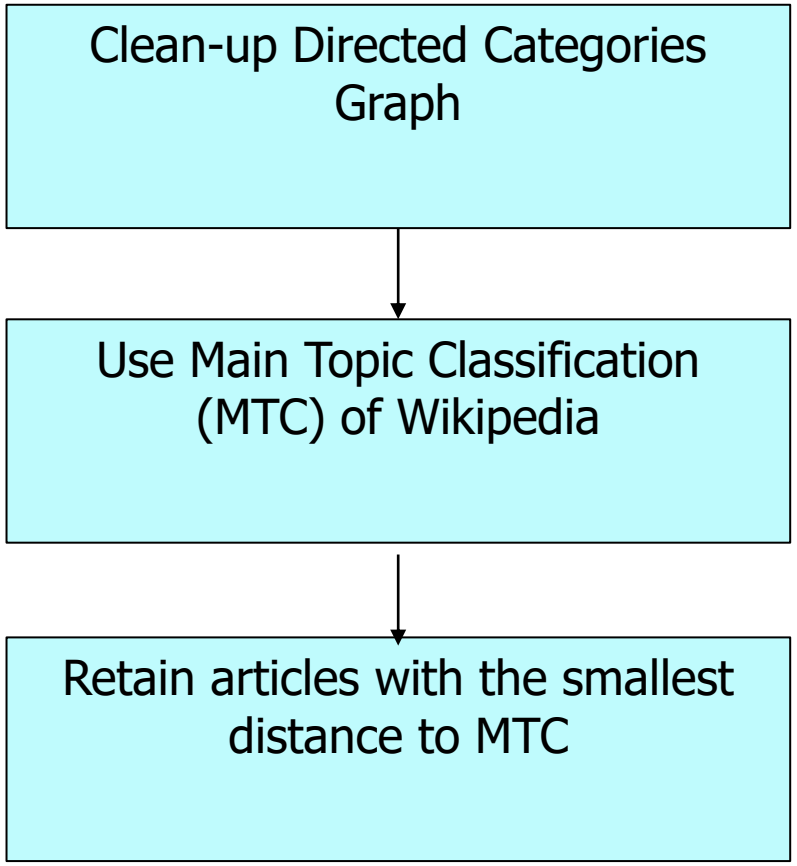
Interest of Public Community:
Wikipedia Pageview Statistics - indicates the number of visitors to a Wikipedia article monthly.

Our Proposed Methodology



- Phase 1: Technology Classification
- Phase 2: Emergence Score

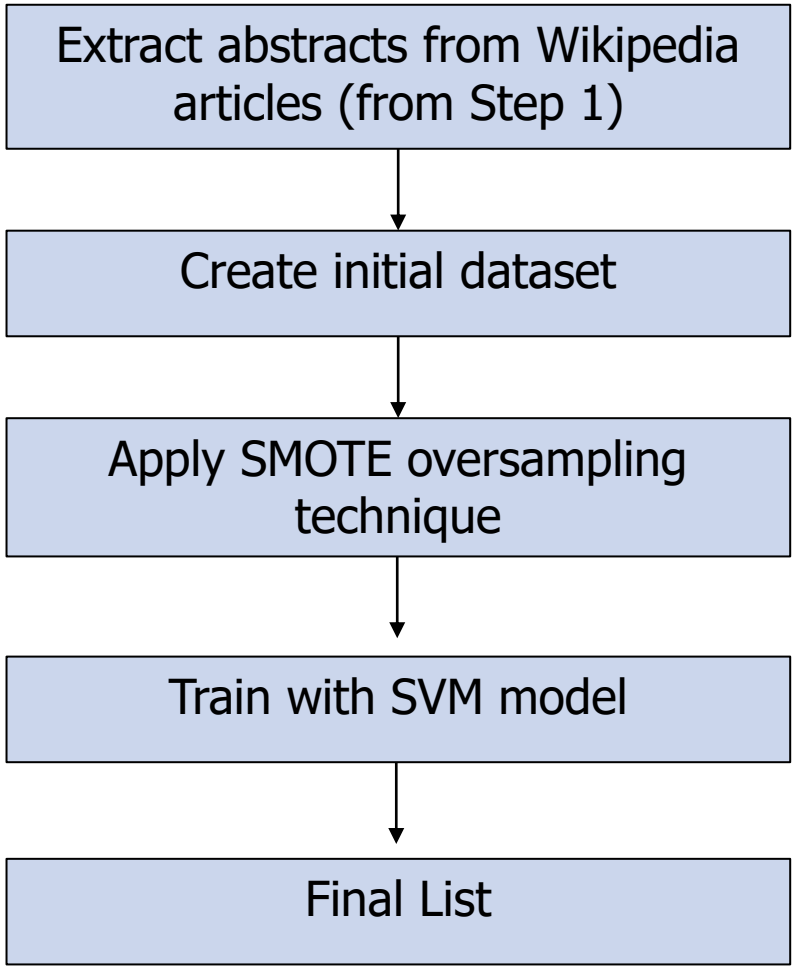
Technology Classification (1)



Algorithm 1 Cleaning and Selecting Relevant Categories

procedure CLEANUPDIRECTEDGRAPH
    Remove hidden categories, admin and user pages from the directed categories graph
    Apply regular expression filters to eliminate irrelevant categories (e.g., companies, people names, brands, currencies, and countries)
end procedure
procedure UTILIZEMAINTOPICCLASSIFICATIONS
    Use Main Topic Classifications (MTC) encompassing categories like Technology, Business, Arts, Health, etc.
    Calculate the shortest path for each category in the filtered graph to MTC
end procedure
procedure FILTERBYDISTANCETOMTC
    Retain articles with the smallest distance to Technology, Science, or Engineering concepts within MTC
end procedure

Technology Classification (2)



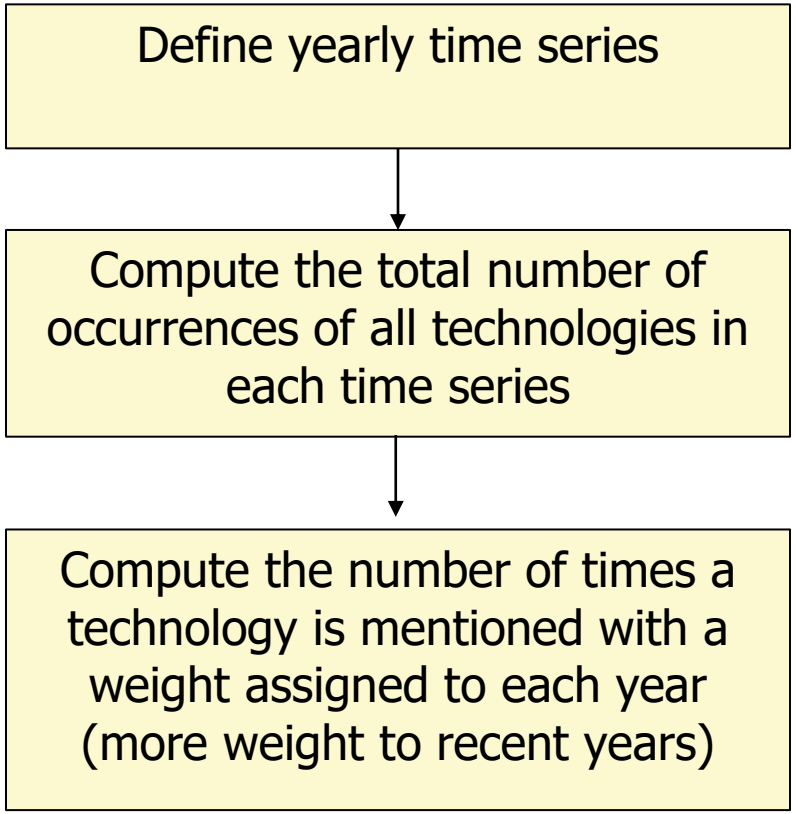
Algorithm 2 Technology Classification using SVM

1: **procedure** CREATEDATASET
2: Extract abstracts from Wikipedia articles in identified technology categories
3: Concatenate and stem abstracts, apply TF-IDF-based weighting
4: Perform feature reduction for usable feature vectors
5: Append distances to each MTC topic to create final feature vectors
6: **end procedure**
7: **procedure** HANDLECLASSIMBALANCE
8: Employ Borderline-SMOTE for oversampling
9: **end procedure**
10: **procedure** FINALIZETECHNOLOGYLIST
11: Use SVM training outcome as the final list of technologies
12: **end procedure**

Emergence Score : Novelty Score

- **Novelty** [def*]: Signifies their distinctive newness, pioneering concepts, break-through advancements, and creative problem-solving, distinguishing them from existing solutions and suggesting transformative potential.

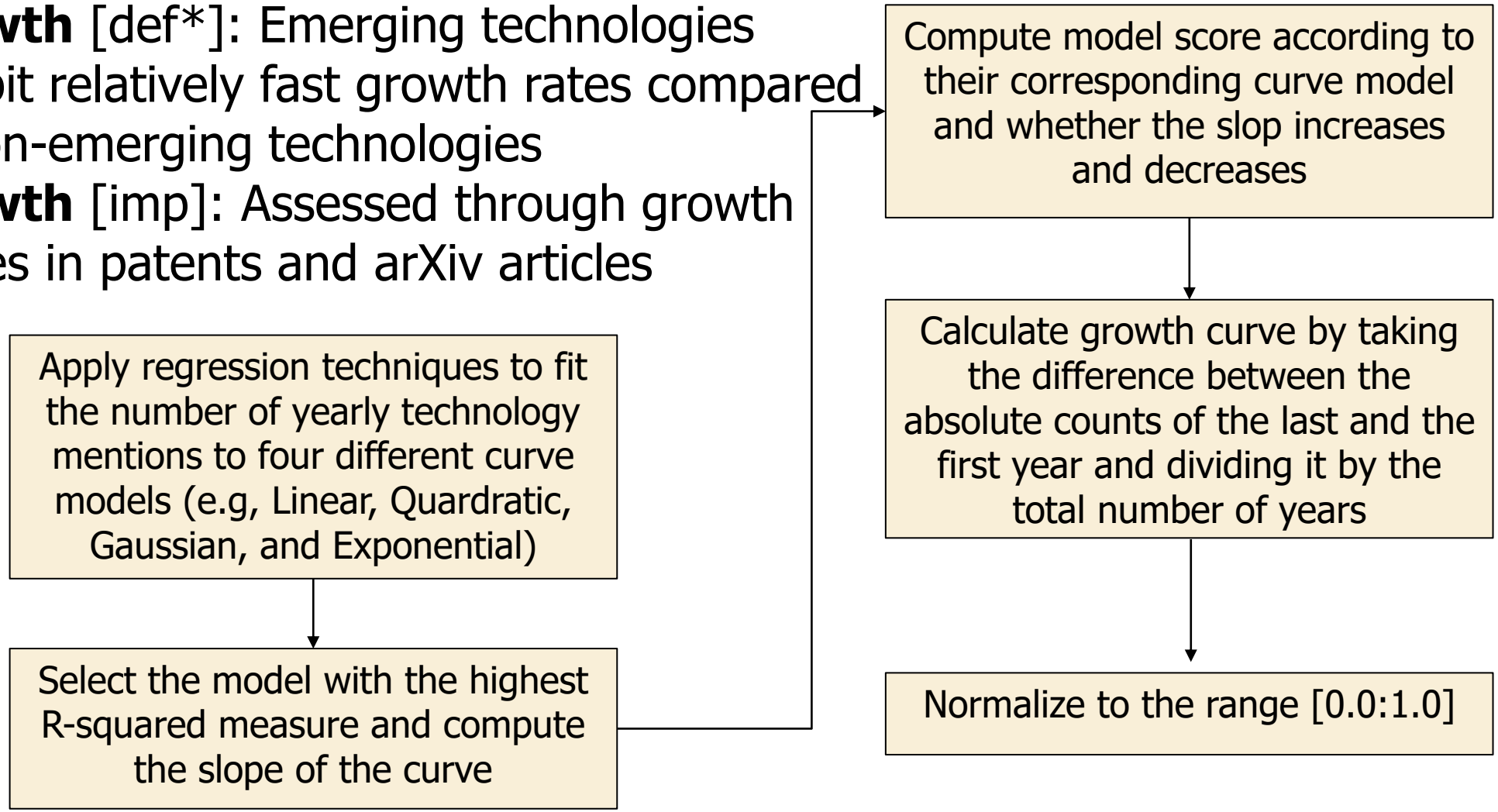
- **Novelty** [imp]: increased mentions in recent years through patents and arXiv articles (i.e., if a particular technology has a significant portion of references occurring in the last few year.)



*D. Rotolo, D. Hicks, B. R. Martin, What is an emerging technology?, Research policy 44 (2015) 1827–1843

Emergence Score : Growth Score

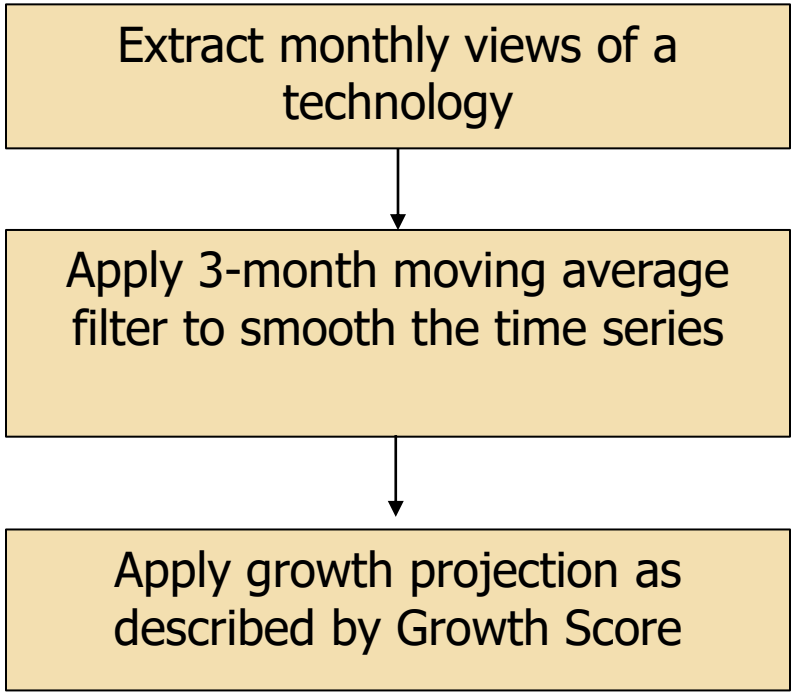
- **Growth** [def*]: Emerging technologies exhibit relatively fast growth rates compared to non-emerging technologies
- **Growth** [imp]: Assessed through growth curves in patents and arXiv articles



*D. Rotolo, D. Hicks, B. R. Martin, What is an emerging technology?, Research policy 44 (2015) 1827–1843

Emergence Score : Impact Score

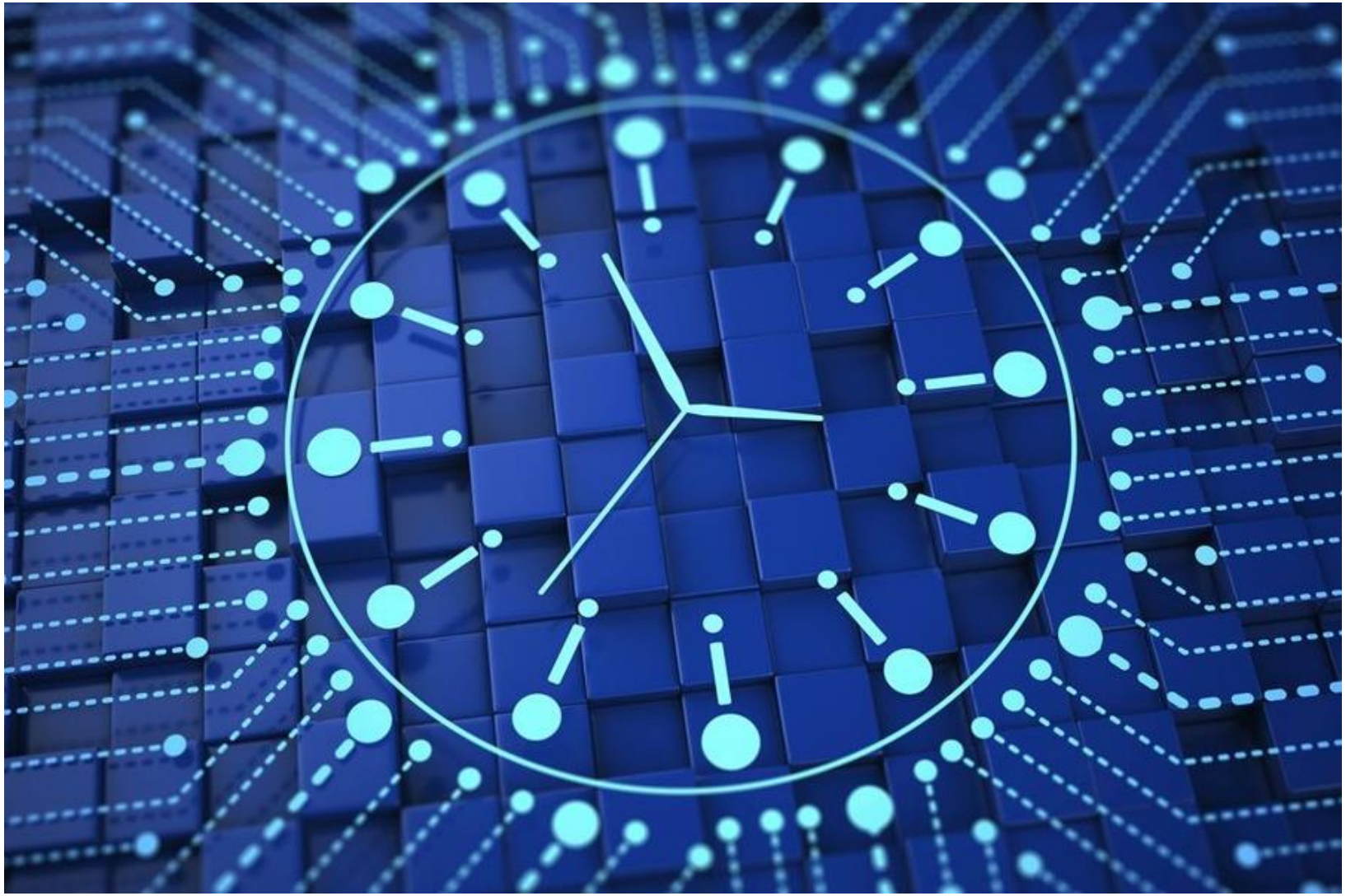
- **Impact** [def*]: More people pay attention to the technology.
- **Impact** [imp]: Wikipedia Pageviews represent the number of times a particular article has been accessed on the Wikipedia website, providing insights into the level of public interest and engagement with specific topics or content.



*D. Rotolo, D. Hicks, B. R. Martin, What is an emerging technology?, Research policy 44 (2015) 1827–1843

Emergence Score : Coherence Score

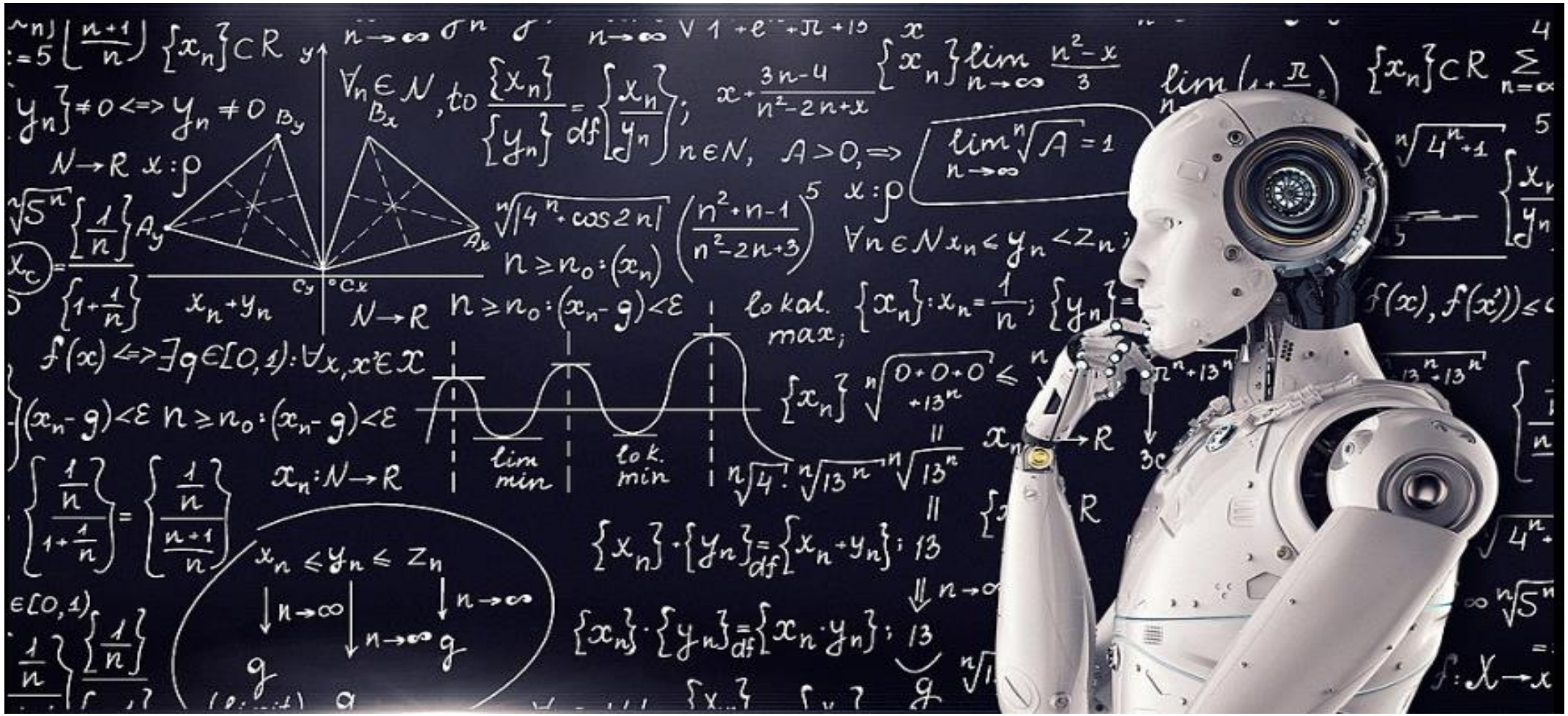
- **Coherence** [def*]: The persistence of a technology over time.
- **Coherence** [imp]: Assume that the presence of a technology on Wikipedia



*D. Rotolo, D. Hicks, B. R. Martin, What is an emerging technology?, Research policy 44 (2015) 1827–1843

Emergence Score

= Novelty Score + Growth Score + Impact Score + Coherence



Top 20 Technologies in Novelty, Growth, and Impact

Novelty	Growth	Impact
Smart City	Communication	URL
Deep Learning†	Wireless	LED Lamp
POWER8	Pixel	Machine Learning†
Vehicle To Everything	Web Server	Artificial Neural Network†
Data Science	Convolutional Neural Network†	Neural Coding
Knowledge Graph	Data Transmission	Robot Locomotion
Internet of Things	Mathematical Optimization	HTTP Cookie
Return-Oriented Programming	Stator	Blockchain
Smartwatch	Rechargeable Battery	Artificial Intelligence†
Multirotor	Radio-Frequency Identification	Computer Science
Ransomware	Unmanned Aerial Vehicle	Sustainable Energy
Row Hammer	Internet of things	BNC Connector
Software-Defined Networking	Quantum Computing	Electron Backscatter Diffraction
Convolutional Neural Network†	Computer Data Storage	Slurry Pump
Virtual Reality Headset	Object Detection	Cryptocurrency
High Efficiency	Video Coding	Lidar Precision and Recall
Cyber-Physical System	Transfer Learning†	XLR Connector
Insider Threat	Unsupervised Learning†	Phishing
Autonomous Car	HVAC	QR Code
Nanosheet	Autonomous Car	PDF

Overall Top 10 Technologies and Technology Classes

Technology
Deep Learning†
Autonomous Car
Internet of Things
Convolutional Neural Network (CNN)†
Machine Learning†
Ransomware*
Key-Value Database
Shard (Database Architecture)
Cyberattack*
Knowledge Graph
Augmented Reality
Smartphone
Communication
Side-Channel Attack*
Cloud Gaming
5G
Data Science
Return Oriented Programming
Lidar
Push Technology

Technology Classes
Artificial Intelligence
Autonomous Driving
Internet of Thing
Computer Security
Database
Knowledge Graph
Augmented, Virtual, Mixed Reality
Connectivity
Telecommunication
Cloud and Virtualization
Data Science
Optical Instrument
Virtual Assistant
Exoskeleton
Computer Vision
Satellite Imagery
Heterogeneous Computing
Distributed Computing
Medical Device
3D Printing

Benchmarking Compatibility

- Compiled the union set of emerging technologies identified by leading technology analysts:
  - ✓ Gartner predicted 35 technologies in its technology hype cycle,
  - ✓ Forrester predicted 12,
  - ✓ IHS Markit 8,
  - ✓ and WEF 10 emerging technologies.
- Upon merging the overlapping technologies from these four lists, we derived a consolidated list of 36 unique technology classes which we use as ground truth.

Technology Classes
Tissue Engineering
Unmanned Aerial Vehicle
Smartdust
Artificial Intelligence
4D Printing⋆
Ontology (Information Science)
Neuromorphic Engineering
Exoskeleton
Edge Computing
Autonomous Driving
Self-Healing System Technology†
Volumetric Display
5G
Quantum Computing
Platform as a Service⋆
Application Specific Integrated Circuits
Autonomous Robot
Mobile Robot
Brain Computer Interface
Internet of Things
Biochip
Digital Twin
Nanotechnology
Virtual Assistant
Lithium-Silicon Battery
Blockchain
Augmented, Virtual, Mixed Reality
E-textiles
Cloud Computing
Computer Vision
Ubiquitous Video†
Natural Language Generation⋆
Switched Fabric
Personalized Medicine
Cell Encapsulation
Gene drive

Table 6
Average Precision (AP) and Recall (R) of Technologies (T) and Technology Classes (TC)

Parameters	Classes	AP	R
base	T	0.72	0.16
max_prec	T	0.81	0.19
	TC	0.72	0.28
	CS TC	0.79	0.36
max_prec_cs	CS TC	0.90	0.36

Conclusion

- Demonstrats automated identification of emerging technologies.
- Time series data analysis on **patents and scientific publications** to capture **novelty** and **growth** scores.
- **Wikipedia pageview statistics** to capture **impact** score.
- **Wikipedia entry** is used to capture **coherence** score.
- **Emergence score** is the sum of **novelty + growth + impact + cohérence**
- > 80% precision compared to the prediction of leading market research firms

Thank You

Any Questions?