

Quarterly Newsletter

CHAIR  
**iDANAE**

INTELLIGENCE · DATA · ANALYSIS · STRATEGY

1Q24

Towards a sustainable  
Artificial Intelligence



POLITÉCNICA

UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

**MS** Management  
Solutions  
*Making things happen*

# Introduction

Artificial Intelligence (AI) has emerged as a transformative force across various sectors, promising to redefine human lives, work, and interactions. However, the pervasive deployment of AI necessitates significant resource consumption, often entailing the utilization of extensive computational infrastructure and advanced hardware accelerators. This, coupled with the training of sophisticated algorithms on large, diverse datasets, may precipitate substantial electricity consumption. Given the variance in energy mixes across nations, the prevalent practices in electricity generation and distribution contribute to the emission of greenhouse gas emissions (GHG) to the atmosphere. Consequently, it is important to assess the implications of AI's development and application on the carbon footprint of both individuals and enterprises, highlighting the environmental impact of technological advancement in this domain.

The computational capacity required for the development and commercialization of more advanced machine learning models has doubled approximately every five to six months since 2010, increasing energy consumption [1]. Recent studies have unveiled that the training of some AI models can result in carbon dioxide emissions equivalent to approximately 284 metric tons, a figure comparable to more than 41 round-trip

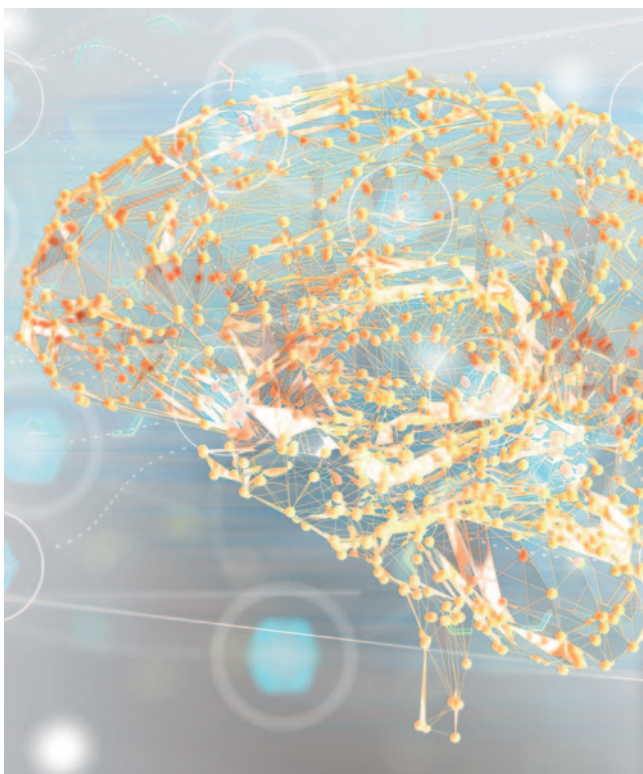
flights between New York City and Sydney, Australia. This revelation brings to light the significant energy consumption and subsequent carbon emissions associated with not only the training of these complex models but also their inference operations (i.e., the real-time processing and analysis of data to provide insights), which continuously accrues as AI services are scaled globally to meet user demands [2].

Given the rapid adoption of AI and its integration into core business functions, the imperative for business executives is to navigate this dual challenge with acumen. The pursuit of innovation should be balanced with a commitment to sustainability, ensuring that the deployment of AI technologies aligns with broader corporate social responsibility goals and environmental stewardship.

All these needs give rise to the concept of green algorithms. Schwartz et al. [3] introduced the term Green AI to refer to "AI research that yields novel results while taking into account the computational cost". This process of making AI development more sustainable is emerging as a possible solution to the problem of the high-power consuming algorithms necessary to train and use AI models. Green AI algorithms are therefore designed to minimize the environmental impact of computing, particularly by reducing the energy consumption and carbon emissions associated with AI models and their training processes. These algorithms aim to achieve efficient computing through various strategies, such as optimizing algorithmic efficiency, compressing the models or employing less computationally intensive models, without compromising on performance.

In addition, from a regulatory perspective the European Union has also taken steps to address these sustainability challenges posed by AI through the recently approved AI Act regulation [4]. It is set to become one of the most comprehensive regulatory frameworks for AI globally to ensure that AI systems are developed and used in a way that is safe, ethical, and respects fundamental rights, including environmental sustainability [5]. This is therefore an additional step towards the greening of the development and use of AI.

In order to reflect on these ideas, this whitepaper aims to analyse possible approaches to sustainable AI, and then discuss two practical use cases on how the IA can contribute to a more sustainable economy.





# An approach to sustainable AI

The environmental impact of AI, and particularly of Generative AI (GenAI), is a growing concern as these technologies are increasingly integrated in business and individuals' life, since the lifecycle of the systems based on these models show a high energy consumption. For example, training Large Language Models (LLM) requires vast amounts of data and multiple training iterations to estimate the billions of parameters present in the model. This leads to significant electricity consumption and large carbon emissions, especially in models based on transformer architectures such as OpenAI GPT-4, Google Gemini or Anthropic Claude, among many others. Both electricity consumption, which evaluates energy efficiency through KWh, and CO<sub>2</sub>-equivalents that measures tons of carbon emissions, are the most common metrics to measure the environmental impact of AI models. For example, GPT3 using 175 billion of parameters emitted x100 times the emissions of an individual in one year of his/her life (see figure 1).

However, the energy consumption used for training the model is often surpassed by the inference phase when millions of users start sending requests to the LLM trained model [7]. The memory resources, the processing power needed for executing these models, and the infrastructure maintenance contribute to the overall energy consumption.

Although complete information (time and resources used) of compute required for inference is not available, in N. Maslej et al [6] benchmarking results of inference performance and energy consumption are presented for the small (7 / 13 billion parameters) and large versions (65 billion parameters) of the Llama LLM developed by Meta. Using different memory and GPUs set ups, the obtained results highlight that even though high-end hardware (like V100 and A100 GPUs) is employed, a considerable portion of GPU memory remains underutilized, suggesting an opportunity for more efficient resource sharing and GPU power capping, which could significantly reduce energy consumption during inference tasks.

In the pursuit of sustainable or Green AI, various strategies have been explored to minimize the environmental impact of AI operations. These strategies include optimizing AI model deployment and hardware to enhance computational efficiency and reduce energy consumption, employing model compression techniques like quantization and pruning to make AI models smaller and faster, developing green algorithms that require less computational power, transitioning AI operations to renewable energy sources to decrease carbon footprints, and implementing dynamic scaling and load management to optimize resource utilization. Together, these approaches offer

Figure 1: Comparison of carbon emissions for different LLMs and real-life examples. Artificial Intelligence Index Report 2023 [6]

## CO<sub>2</sub> Equivalent Emissions (Tonnes) by Selected Machine Learning Models and Real Life Examples, 2022

Source: Luccioni et al., 2022; Strubell et al., 2019 | Chart: 2023 AI Index Report

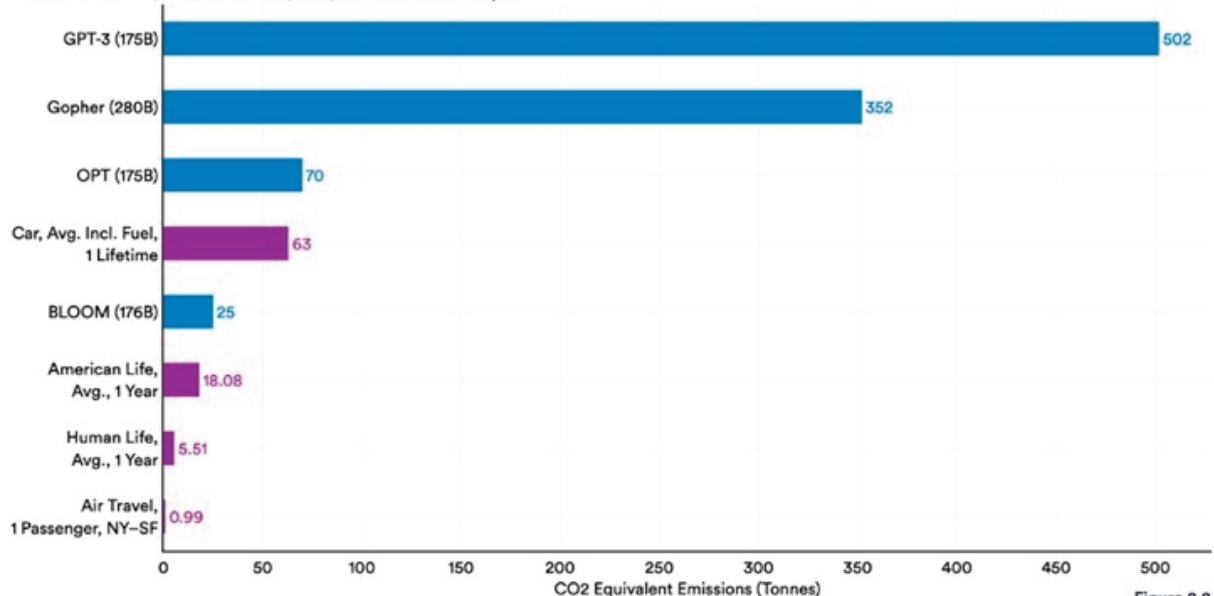


Figure 2.8.2

a multi-faceted roadmap towards more energy-efficient and environmentally responsible AI systems, addressing both the urgent need for sustainability and the continuous demand for advanced AI capabilities [8][9][10]:

- ▶ **Optimized model deployment and hardware:** Utilizing energy-efficient hardware like GPUs and TPUs, specifically designed for AI tasks, significantly enhances computational efficiency. This strategy not only accelerates the processing speed but also minimizes energy consumption during AI model training and inference. By optimizing software to fully leverage the capabilities of these specialized hardware units, organizations can achieve substantial reductions in energy use, contributing to more sustainable AI operations.
- ▶ **Model compression techniques:** Implementing model compression techniques such as quantization, which reduces the precision of number representations in arithmetic computations, and pruning, which eliminates unnecessary weights in the AI model, can significantly decrease both the size of AI models and their computational requirements. These methods enhance the speed of AI models while also lowering energy consumption, making AI systems more efficient and environmentally friendly.
- ▶ **Green algorithms:** Focusing on developing algorithms that require less computational power without compromising effectiveness, green algorithms aim to reduce the environmental impact of AI. This involves innovative research into creating AI models and algorithms that are not only powerful but also designed with energy efficiency in mind, addressing the dual challenge of maintaining high performance while minimizing carbon emissions.

- ▶ **Renewable energy sources:** Transitioning data centers that power AI operations to renewable energy sources is a critical step in reducing the carbon footprint of AI activities. By sourcing energy from renewable resources for both training and inference phases of AI models, companies can significantly lower the environmental impact of their AI operations, aligning technology advancement with sustainability goals. It is important that this transition does not translate into increased water consumption for cooling.
- ▶ **Dynamic scaling and load management:** By dynamically adjusting computational resources based on current demand, this strategy ensures that energy is not wasted during low usage periods. This approach to resource allocation not only conserves energy but also optimizes the overall efficiency of AI systems, leading to a more sustainable utilization of computational resources across different scenarios.

While exploring strategies for more energy-efficient AI, it is also important to consider the regulatory landscape shaping the development and deployment of these technologies and promoting sustainable practices. The European Union's AI Act, which is set to become one of the most comprehensive regulatory frameworks for AI globally, addresses this issue by requesting from the providers of general-purpose models (e.g. Large Language Models) design specifications of the model and training process, information on the data used for training, testing and validation, computational resources used to train the model as well as reporting known or estimated energy consumption of the model based on information about computational resources used (AI Act [4], Article 53(1), point (a)). This sets a guideline and a first approach to create awareness and starts putting some control practices to the organizations that implement and deliver AI systems to be used at global scale.



# Practical applications

As discussed in the previous section, the rapid application of artificial intelligence needs ongoing discussions about its sustainability and environmental impact. As AI technologies, and particularly LLM, become integral in processing and analyzing vast amounts of data, their energy and computational demands have surged. Moreover, besides LLM, there are other AI technologies that also process huge amounts of data, such as those captured through mobile phones and from sensor networks. Although each of these individual AI models may consume a small amount of power (both standalone and when compared to LLM), the large number of those models that could be installed in many different devices requires special attention. In addition, this problem is aggravated by the fact that new data are frequently collected, allowing a more frequent retraining. Recognizing all these challenges, two innovative use cases are presented to discuss these concerns and show how sustainability practices could be integrated in AI across industries.

## Use case 1: GreenLightningAI

There has been a lot of attention into how energy consumption is distributed throughout the training process of AI models. However, since data changes over time, new patterns in the data are necessary to be recognized and AI models need to be refreshed accordingly to preserve model's accuracy. As the demand and complexity of AI applications soar, even with advanced algorithmic, mathematical, and hardware improvements, the process of training complex Deep Neural Networks (DNN) remains economically and environmentally expensive. Traditional optimizations for DNN training are hitting limits, prompting a need for a fundamentally different approach. GreenLightningAI [11] is a solution developed by Qsimov Quantum Computing in collaboration with Universitat Jaume I de Castelló and Universitat Politècnica de València. It has been fully implemented, tested, and evaluated by Management Solutions in a project directed by Qsimov Quantum Computing. It is an innovative AI system designed to address the escalating computational and environmental costs associated with training DNN. More specifically, the objective of the authors is to simplify, accelerate, and make re-training of AI models more environmentally efficient.







The proposed solution includes decoupling structural and quantitative knowledge in AI systems. It employs a linear model to emulate the piece-wise linear behavior of DNN. This model subsets itself for each specific sample, storing the information required for subset selection (structural information) separately from the linear model parameters (quantitative knowledge). Since structural information is much more stable than quantitative knowledge, only the linear model parameters need to be re-trained. This innovative design allows for easy combination of multiple model copies trained on different datasets, facilitating faster and greener re-training algorithms, including incremental re-training and federated incremental re-training.

Structural knowledge is represented by the activation patterns in the neural network for the different samples. Each activation pattern consists of a set of active paths. The stability of the structural knowledge is attributed to the early capture of high-level features in training, indicating that these high-level features have been effectively learned. The system is defined using two elements:

- ▶ Path selector: which identifies the active neuron paths for each individual sample and remains unchanged during re-training.
- ▶ Estimator: which contains the linear model and uses the path selector's outputs for subsetting said linear model for each individual sample both for inference and (re)-training. The linear model parameters are updated during (re)-training.

Among the benefits of this system design, it is worth noting that the estimator consists of just one hidden layer, which allows a much faster training and avoid the problems related to vanishing gradient. In addition, since a linear system is used, the incremental retraining is allowed, resulting in an expected sharp reduction of retraining times. It also provides direct interpretability.

This new promising training method validates the potential of green algorithms in AI, aligning with environmental sustainability goals by optimizing training processes for lower energy consumption. The experiments validate the hypothesis that structural information stabilizes earlier than quantitative knowledge and can be kept unmodified during re-training, while achieving comparable validation accuracy with traditional re-training methods but with much reduced resource usage.

### ***Use case 2: Interpretable Machine Learning to Estimate Corporate Emissions***

Artificial Intelligence needs to be analysed from a sustainability perspective, but it can also be used as a tool to help organizations identify and close sustainability gaps. One of the gaps that some companies have is reporting Scope 1 and Scope 2 GHG<sup>1</sup> because

<sup>1</sup>Scope 1 refers to direct GHG emissions covered by the Kyoto Protocol produced by sources that are owned or controlled by a company. For example, emissions from combustion in owned or controlled boilers, furnaces, vehicles, etc.; emissions from chemical production in owned or controlled process equipment. Scope 2 accounts for GHG emissions from the generation of purchased electricity consumed by the company. For further information, refer to [12].



on some occasions such information is not easy to calculate. The sustainability reports for these emissions are sometimes estimated as not all companies have accurate data on emissions. Even regulatory entities are aware of how complicated this task can be, so they allow the use of estimations or calculations based on best available information when data are not available.

A recent study [13] introduces a machine learning model to estimate corporate Scope 1 and Scope 2 greenhouse gas emissions designed to be transparent and interpretable. It aims to provide accurate emissions estimates across different sectors, countries, and revenue sizes, contributing to a more standardized approach to sustainability reporting.

The system employs regression settings, separating scope 1 and scope 2 emissions into two models. These models take features that provide a comprehensive view of a company's operations, financial health, and environmental impact. Features are divided into 4 groups:

- ▶ General features: Year, Country.
- ▶ Industry Classification: BICS Classification Levels 1 to 7, New Energy Exposure Rating.
- ▶ Financial features: Employees, Capital Expenditure, Enterprise Value, Revenues, Property, Plant & Equipment (Gross and Net), Depreciation, Depletion & Amortization, Energy Consumption, Total Power Generated.

- ▶ Regional Features: Country Energy Mix Carbon Intensity, Existence of an Emission Trading System (ETS) or carbon taxes

The target GHG emissions are obtained from high-quality sources like CDP and Bloomberg, prioritizing audited and non-modeled emissions. The data undergoes a cleaning process to address variability and inconsistencies, with emissions reported at different times being aligned to a consistent annual timeframe. The system then divides the dataset into training, validation, and test sets, optimizing the model to minimize mean-squared error. Gradient Boosting Decision Trees (GBDT) like LightGBM are utilized due to their ability to learn more generic functional forms and handle non-linear patterns. The model is trained to estimate the decimal logarithm of GHG emissions to handle the skewed nature of emissions data effectively.

The interpretability of this model resides in the use of Shapley Values a concept borrowed from game theory and applied to machine learning to assess the contribution of each feature towards the model's predictions. Shapley values provide a systematic way to determine how each input feature influences the output of the model, thereby offering insights into the decision-making process of the machine learning model. These values provide a detailed explanation of the prediction, making the model's decisions transparent and understandable, which is particularly valuable in assessing the importance of variables in complex models like those estimating greenhouse gas emissions. It also addresses a common criticism of Gradient Boosted Decision Trees models, which are known for their superior performance in handling tabular data but criticized for their lack of interpretability.

According to the results presented, the model shows good global performance, with stable results across different sectors, countries, and revenue deciles. Specifically, the model demonstrates particularly good performance in the most emissive sectors, highlighting the effectiveness of the chosen sectorization methodology and the model's ability to adapt to the complex variability inherent in GHG emissions data. Moreover, the system's approach to training and validation ensures that the model is well-equipped to handle the latest and most relevant data, further enhancing its accuracy and reliability in estimating GHG emissions across a wide range of companies and industries.





# Conclusions

The significant environmental impact of Artificial Intelligence, particularly the substantial energy consumption and carbon emissions from both the development and application of large language models, highlights a need for a shift towards Green AI practices. These practices, which include optimized model deployment, model compression, development of energy-efficient algorithms, adoption of renewable energy sources, and dynamic resource scaling, represent a holistic approach towards reducing AI's carbon footprint.

Through practical applications, such as innovative methods for AI model re-training and machine learning models for estimating corporate emissions, the potential of AI as a tool for enhancing environmental sustainability has been illustrated. These examples demonstrate that integrating sustainability into AI development not only mitigates its environmental impact but also leverages AI to support broader sustainability efforts across various sectors.

Regulatory frameworks, such as the European Union's AI Act, are going to play an important role in promoting sustainable practices within the AI industry by setting standards for energy consumption transparency and the implementation of environmentally friendly technologies.

In summary, balancing AI innovation with environmental stewardship is of paramount importance. By embracing Green AI practices, adhering to regulatory standards, and exploring sustainable applications, progress can be made towards a future where AI contributes positively to the environmental objectives, ensuring that technological advancement progresses in harmony with sustainability.





# Bibliography

- [1] International Energy Agency, 'Why AI and Energy Are the New Power Couple', [Online]. Available: <https://www.iea.org/commentaries/why-ai-and-energy-are-the-new-power-couple>.
- [2] Strubell, E., Ganesh, A., & McCallum, A. (2019). "Energy and Policy Considerations for Deep Learning in NLP." arXiv preprint arXiv:1906.02243.
- [3] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Communications of the ACM*, vol. 63, no. 12, pp. 54-63, Dec. 2020. [Online]. Available: DOI:10.1145/3381831.
- [4] European Parliament, Artificial Intelligence Act. European Parliament legislative resolution of 13 March 2024 P9\_TA(2024)0138. [Online]. Available: [https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf).
- [5] Artificial Intelligence Act: MEPs adopt landmark law. Press release 13.03.2024. Retrieved from <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>
- [6] N. Maslej, L. Fattorini, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, H. Ngo, J. C. Niebles, V. Parli, Y. Shoham, R. Wald, J. Clark, and R. Perrault, "The AI Index 2023 Annual Report," AI Index Steering Committee, Stanford Institute for Human-Centered AI (HAI), Stanford University, Stanford, CA, Apr. 2023.
- [7] J. McDonald, B. Li, N. Frey, D. Tiwari, V. Gadepally, S. Sams: Great Power, Great Responsibility: Recommendations for Reducing Energy for Training Language Models. MIT Lincoln Laboratory and Northwestern University. May 2022. arXiv:2205.09646
- [8] "Analog AI Chip for Low Power," IBM Research Blog. [Online]. Available: <https://research.ibm.com/blog/analog-ai-chip-low-power>.
- [9] "AI Hardware for Hybrid Cloud," IBM Research Blog. [Online]. Available: <https://research.ibm.com/blog/ai-hardware-hybridcloud>.
- [10] J. F. Duato Marín, "Hacia una inteligencia artificial medioambientalmente sostenible," presented at the Real Academia de Ciencias Exactas, Físicas y Naturales de España, Madrid, Spain, Sep. 21, 2023.
- [11] J. Duato, J. I. Mestre, M. F. Dolz, and E. S. Quintana-Ortí, "GreenLightningAI: An Efficient AI System with Decoupled Structural and Quantitative Knowledge," arXiv:2312.09971v1 [cs.LG], 15 Dec. 2023.
- [12] The Greenhouse Gas Protocol: A Corporate Accounting and Reporting Standard. Revised edition.
- [13] J. Assael, T. Heurtebize, L. Carlier, and F. Soupé, "Greenhouse Gases Emissions: Estimating Corporate Non-Reported Emissions Using Interpretable Machine Learning," *Sustainability*, vol. 15, no. 3391, Feb. 2023. [Online]. Available: <https://doi.org/10.3390/su15043391>



#### Other references of interest

- [14] S. Samsi, D. Zhao, J. McDonald, B. Li, A. Michaleas, M. Jones, W. Bergeron, J. Kepner, D. Tiwari, and V. Gadepally, "From Words to Watts: Benchmarking the Energy Costs of Large Language Model Inference," arXiv:2310.03003v1 [cs.CL], 4 Oct. 2023.
- [15] Google. (2021). "Google Environmental Report." <https://sustainability.google/reports/environmental-report-2021/>
- [16] Amazon. (2020). "Amazon Sustainability." <https://sustainability.aboutamazon.com/>
- [17] Microsoft. (2021). "Microsoft Environmental Sustainability Report." <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4w6Y7>
- [18] S. A. Khowaja, P. Khuwaja, and K. Dev, "ChatGPT Needs SPADE (Sustainability, Privacy, Digital divide and Ethics) Evaluation: A Review," University of Sindh, Pakistan, and Munster Technological University, Ireland, arXiv:2305.03123v1 [cs.CY], 13 Apr. 2023.
- [19] "The Green Algorithm: Measuring Sustainability in AI," Version 1, Medium. [Online]. Available: <https://medium.com/version-1/the-green-algorithm-measuring-sustainability-in-ai-4c775e811c2a>.
- [20] R. Verdecchia, J. Sallou, and L. Cruz, "A Systematic Review of Green AI," arXiv:2301.11047v3 [cs.AI], 5 May 2023.
- [21] C.-J. Wu et al., "Sustainable AI: Environmental Implications, Challenges, and Opportunities," arXiv:2111.00364v2 [cs.LG], 9 Jan. 2022.
- [22] Williams, J. (2023). "AI and sustainability: Balancing innovation with environmental impact," SAS Voices. Accessed at SAS Blogs.
- [23] Nature Climate Change. (2020). "The Carbon Footprint of AI and Cloud Computing." Nature.
- [24] "Achieving a sustainable future for AI," MIT Technology Review, provided by Intel. Accessed at MIT Technology Review.

---

#### Authors

Ernestina Menasalvas (UPM)  
Manuel Ángel Guzmán (Management Solutions)  
Sergio Ruiz (Management Solutions)  
Guillermo Fragio (Management Solutions)







**POLITÉCNICA**

UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

**MSO** *Management  
Solutions*  
*Making things happen*

The Universidad Politécnica de Madrid is a multi-sector and multi-disciplinary Public Law Entity, which carries out teaching, research and scientific and technological development activities.

[www.upm.es](http://www.upm.es)

Management Solutions is an international consulting firm, focused on business, finance, risk, organization, technology and process consulting, operating in more than 50 countries and with a team of close to 4,000 professionals working for more than 1,800 clients worldwide.

[www.managementsolutions.com](http://www.managementsolutions.com)

For more information visit

**[blogs.upm.es/catedra-idanae/](http://blogs.upm.es/catedra-idanae/)**