

## Data Science for Social Good: between the red and the blue pills

« This is your last chance. After this, there is no turning back. You take the *blue* pill—the story ends, you wake up in your bed and believe whatever you want to believe. You take the *red* pill—you stay in Wonderland, and I show you how deep the rabbit hole goes. Remember: all I'm offering is the truth. Nothing more. »  
Morpheus, The Matrix, 1999

Datasets of heterogeneous data have become the backbone of scientific, analytic, and forecasting processes. Datasets represent partial and incomplete observations of phenomena produced within complex systems that can be analysed. Combining simulation techniques, artificial vision, and artificial learning with data science techniques, it is possible to compute mathematical models to understand and predict phenomena. To achieve this, data must go through complex and repetitive pipelines, namely *data science pipelines*. A data science pipeline combines data visualisation, cleaning, preparation, modelling and prediction, and assessment tasks. These tasks define a dataflow intended to explore, model or understand observations to get close to a possible understanding of the intrinsic knowledge within them. The phenomena that determine the world and its citizens destinies can be thereby apprehended, controlled and shaped thanks to data (!).

Beyond the technical and exciting questions and challenges that data science pipelines, the question is whether they can be used for social good.

*"A social good is something that benefits the largest number of people in the largest possible way, such as clean air, clean water, healthcare, and literacy."*

Can data scientists be agents of change? Is change accessible in a democratised manner thanks to data science? Through the glass of technical data science details, this talk will introduce basic concepts, associated knowledge and examples of social good that is done in favour of some communities by optimistic data scientists.

### Short bio

Genoveva Vargas-Solar (<http://www.vargas-solar.com>) is principal scientist of the French Council of Scientific Research (CNRS), France. She is regular member of the Mexican Academia of Computing. From 2008-2020, she was deputy director the Franco-Mexican Laboratory of Informatics and Automatic Control (LAFMIA) an international research unit established at CINVESTAV.

In 2014 she obtained her *Habilitation à Diriger des Recherches* (HDR - tenure) from University of Grenoble. In 2000, she obtained her first PhD degree in Computer Science at University Joseph Fourier and in 2005 she obtained her second PhD degree in Literature at University Stendhal. In 1997 she obtained her first Master Degree in Computer Science at University Joseph Fourier and in 1998 she obtained her second Master Degree in Compared Literature at

University Stendhal. She did her undergraduate studies on Computer Systems Engineering at Universidad de las Américas in Puebla.

Her research interests in Computer Science concern distributed and heterogeneous databases, reflexive systems and service-based database systems. Particularly, she contributes to the construction of service-based database management systems. The objective is to design data management services guided by Service Level Agreements (SLA) to provide methodologies, algorithms and tools for integrating, deploying and executing a service-oriented data management functions. Quality of service criteria and behaviour properties included in SLA are adapted to applications requirements (e.g., transactional execution, reliability, fault tolerance, evolution and dynamic adaptability). She conducts fundamental and applied research activities for addressing these challenges on different architectures ARM, raspberry, cluster, cloud, and HPC. She has applied her results to e-Science applications in Astronomy, Biology, social sciences, digital-humanities, and industry 4.0.

She is an active militant of data and A+I decolonisation and decolonial feminism in data science in the global south. She is a member of the database interconference diversity and inclusion (D&I) initiative on she runs several gender equalities actions in the global north. Her research interests in Literature concern middle age Literature, myths' critics and myths' analysis applied to different myths of origins.

She has coordinated several research projects in Europe and Latin America financed by governments and industrial partners. She actively promotes the scientific cooperation in Computer Science between Latin America and Europe particularly between France and Mexico.

## Ciencia de datos para el bien social: entre las píldoras rojas y las azules

" Esta es tu última oportunidad. Después de esto, no hay vuelta atrás. Si tomas la píldora azul, el cuento termina, te despiertas en tu cama y crees lo que quieras. Si tomas la píldora roja, te quedas en el País de las Maravillas y te enseño hasta dónde llega la madriguera del conejo. Recuerda: todo lo que ofrezco es la verdad. Nada más. "

Morfeo, The Matrix, 1999

Los conjuntos de datos heterogéneos se han convertido en la columna vertebral de los procesos científicos, analíticos y de previsión. Los conjuntos de datos representan observaciones parciales e incompletas de fenómenos producidos en sistemas complejos que pueden ser analizados. Combinando técnicas de simulación, visión artificial y aprendizaje artificial con técnicas de ciencia de datos, es posible calcular modelos matemáticos para comprender y predecir fenómenos. Para lograrlo, los datos deben pasar por procesos complejos y repetitivos, es decir, procesos de ciencia de datos. Un proceso de ciencia de datos combina tareas de visualización, limpieza, preparación, modelización y predicción, y evaluación de datos. Estas tareas definen un flujo de datos destinado a explorar, modelar o comprender las observaciones para acercarse a una posible comprensión del conocimiento intrínseco en ellas. Los fenómenos que determinan el mundo y los destinos de sus ciudadanxs (se lee ciudadanix) pueden así ser aprehendidos, controlados y modelados gracias a los datos (i!).

Más allá de las cuestiones y retos técnicos y apasionantes que plantea la ciencia de los datos, la cuestión es si pueden utilizarse para el bien social.

*"Un bien social es algo que beneficia al mayor número de personas de la mayor manera posible, como el aire limpio, el agua potable, la asistencia sanitaria y la alfabetización".*

¿Pueden lxs (se lee lix) científicxs (se lee científicix) de datos ser agentes de cambio? ¿Es el cambio accesible de forma democratizada gracias a la ciencia de datos? A través del cristal de los detalles técnicos de la ciencia de datos, esta charla introducirá conceptos básicos, conocimientos asociados y ejemplos del bien social que se hace en favor de algunas comunidades por parte de científicxs de datos optimistas.

### Breve biografía

Genoveva Vargas-Solar (<http://www.vargas-solar.com>) es científica principal del Consejo Nacional de Investigación Científica (CNRS), Francia. Es miembrx (se lee miembrix) titular de la Academia Mexicana de Informática. De 2008 a 2020, fue subdirectora del Laboratorio Franco-Mexicano de Informática y Control Automático (LAFMIA) una unidad de investigación internacional establecida en el CINVESTAV.

En 2014 obtuvo su Habilitation à Diriger des Recherches (HDR - tenure) de la Universidad de Grenoble. En el año 2000, obtuvo su primer doctorado en Informática en la Universidad Joseph Fourier y en 2005 obtuvo su segundo doctorado en Literatura en la Universidad Stendhal. En 1997, obtuvo su primer Máster en Informática en la Universidad Joseph Fourier y en 1998 obtuvo su segundo Máster en Literatura Comparada en la Universidad Stendhal. Realizó sus

estudios de licenciatura en Ingeniería de Sistemas Computacionales en la Universidad de las Américas en Puebla.

Sus intereses de investigación en Ciencias de la Computación se refieren a bases de datos distribuidas y heterogéneas, sistemas reflexivos y sistemas de bases de datos basados en servicios. En particular, contribuye a la construcción de sistemas de gestión de bases de datos basados en servicios. Su objetivo es diseñar servicios de gestión de datos guiados por Acuerdos de Nivel de Servicio (SLA) para proporcionar metodologías, algoritmos y herramientas para integrar, desplegar y ejecutar funciones de gestión de datos orientadas al servicio. Los criterios de calidad de servicio y las propiedades de comportamiento incluidas en los SLA se adaptan a los requisitos de las aplicaciones (por ejemplo, ejecución transaccional, fiabilidad, tolerancia a fallos, evolución y adaptabilidad dinámica). Lleva a cabo actividades de investigación fundamental y aplicada para hacer frente a estos retos en diferentes arquitecturas ARM, raspberry, clúster, nube y HPC. Ha aplicado sus resultados a aplicaciones de e-ciencia en astronomía, biología, ciencias sociales, humanidades digitales e industria 4.0.

Es una activa militante de la descolonización de los datos y de los algoritmos en inteligencia artificial y militante del feminismo decolonial en la ciencia de los datos en el sur global. Es miembro de la iniciativa de diversidad e inclusión (D&I) de la interconferencia de bases de datos y dirige varias acciones de igualdad de género en el norte global. Sus intereses de investigación en el campo de la literatura se centran en la literatura de la edad media, la crítica de los mitos y el análisis de los mitos aplicados a diferentes mitos de los orígenes.

Ha coordinado numerosos proyectos de investigación en Europa y América Latina financiados por gobiernos y socios industriales. Promueve activamente la cooperación científica en informática entre América Latina y Europa, en particular entre Francia y México.