Gabriel Henrique **Lopes Gomes Alves Nunes**

PhD Candidate in Computer Science | MSc in Computer Science, BSc in Physics

nunesgh.com nunesgh.com/scholar nunesgh.com/orcid nunesgh.com/lattes

Sydney, New South Wales, Australia

Cotutelle Doctoral candidate in Computer Science at the Federal University of Minas Gerais (UFMG), Brazil, and at Macquarie University, Australia. Master in Computer Science and Bachelor in Physics from UFMG. Interested in Formal Methods, Quantitative Information Flow, Responsible Computing (e.g. Privacy, Utility, Fairness), Artificial Intelligence, and Neuroscience.

Member of Topete Research Group and INSCRYPT | T-Rex Laboratory.

Education

02/2023 – 06/2025 Doctoral Degree in Computer Science with an International Macquarie University Research Excellence (iMQRES) Scholarship at the School of Computing of Macquarie University (Researcher Profile). Supervised under Cotutelle by Prof. Annabelle McIver.

07/2021 – 06/2025 Doctoral Degree in Computer Science with a CAPES Scholarship at the Graduate Program in Computer Science of UFMG. Supervised under Cotutelle by Prof. Mário Alvim.

- Research area: Quantitative Information Flow theory and its application to precisely characterize how different methods of noise introduction affect the trade-off between privacy and utility in microdata and statistical data publications.

03/2019 – 04/2021 Master Degree in Computer Science with a CNPq Scholarship at the Graduate Program in Computer Science of UFMG. Supervised by Prof. Mário Alvim and by Prof. Annabelle McIver.

- Dissertation defended and approved in April 28, 2021: A formal quantitative study of privacy in the publication of official educational censuses in Brazil (**HDL**: 1843/38085).
- Dissertation selected for the 35th Thesis and Dissertation Contest (CTD 2022), as part of the 42nd Congress of the Brazilian Computing Society (CSBC 2022) (**DOI**: 10.5753/ctd.2022.223158).

03/2014 – 07/2018 Bachelor Degree in Physics at UFMG.

03/2012 – 12/2013 Bachelor Degree in Nanotechnology at Federal University of Rio de Janeiro. Unfinished.

02/2011 – 06/2011 Bachelor Degree in Medicine at Medical Sciences of Minas Gerais. Unfinished.

Experience

07/2023 – 06/2025 | **Sessional Teaching Academic**, Macquarie University, Sydney, New South Wales, Australia

Offensive Security (COMP2320/COMP6320), Data Privacy and Information Security (COMP3300), Secure Applications Development (COMP3310), and Formal Methods (COMP4000) at the School of Computing.

Computer Science | Ethical Hacking | Formal Methods | Java | Privacy | Python | Quantitative Information Flow | Security

07/2022 – 11/2022 | **Student Researcher**, Google LLC, New York, New York, USA

Internship supervised by Dr. Andrés Muñoz Medina.

Computer Science | Apache Beam | Machine Learning | Quantitative Information Flow | Privacy | Python

12/2020 – 03/2021 | **Information Security Analyst**, Research Development Foundation, UFMG, Belo Horizonte, Minas Gerais, Brazil

Information Security Analyst in the project PRICE - Privacy in Educational Censuses.

Computer Science | Privacy | Python | Quantitative Information Flow | Transparency

09/2019 – 11/2019 | **Visiting Scholar**, Macquarie University, Sydney, New South Wales, Australia

Internship at the Department of Computing. Supervised by Prof. Annabelle McIver.

Computer Science | Privacy | Python | Quantitative Information Flow | Transparency

03/2019 – Present | **Volunteer IT System Administrator**, UFMG, Belo Horizonte, Minas Gerais, Brazil

INSCRYPT | T-Rex Laboratory.

Computer Science | Linux | Debian | System Administration | Volunteering

01/2016 – 10/2014 | **Undergraduate Tutoring & Technological and Industrial Initiation**, Room of Physics Demonstrations, UFMG, Belo Horizonte, Minas Gerais, Brazil

- Development of a high-resolution and low-cost optical spectrometer and of an electrical Paul's Trap.
- PROGRAD/UFMG & CNPq Scholarships. Supervised by Prof. Elmo Salomão.

Physics | Spectrometer | LabVIEW | SolidWorks

Awards

Excellence in Higher Degree Research (Student Award) — December 2024
Faculty of Science and Engineering, Macquarie University.

Outstanding Poster — June 2024
Future Communications Research Centre, Macquarie University.
Poster: Quantitative Information Flow for Privacy Analysis (Future Communications Research Centre Workshop).
Computer Science | Topics API | Third-Party Cookies | Quantitative Information Flow | Privacy | Utility | Privacy-Utility Trade-Off

Research Rising Star — October 2023
School of Computing, Faculty of Science and Engineering, Macquarie University.
Paper: A Novel Analysis of Utility in Privacy Pipelines, Using Kronecker Products and Quantitative Information Flow.
Computer Science | Privacy | Privacy and Utility Trade-off | Quantitative Information Flow

Google LARA Research Scholarship — February 2022
9^{th} Google Latin America Research Awards (LARA).
A robust and explainable QIF-based framework for assessing big data privacy risks.
Computer Science | Quantitative Information Flow | Disclosure Control | Microdata | Differential Privacy | Privacy

Conference Papers

The Privacy-Utility Trade-off in the Topics API — October 2024
2024 ACM SIGSAC Conference on Computer and Communications Security. **DOI**: 10.1145/3658644.3670368. **arXiv**: 2406.15309.
The ongoing deprecation of third-party cookies by web browser vendors has sparked the proposal of alternative methods to support more privacy-preserving personalized advertising on web browsers and applications. The Topics API is being proposed by Google to provide third-parties with "coarse-grained advertising topics that the page visitor might currently be interested in". In this paper, we analyze the re-identification risks for individual Internet users and the utility provided to advertising companies by the Topics API, i.e. learning the most popular topics and distinguishing between real and random topics. We provide theoretical results dependent only on the API parameters that can be readily applied to evaluate the privacy and utility implications of future API updates, including novel general upper-bounds that account for adversaries with access to unknown, arbitrary side information, the value of the differential privacy parameter ϵ, and experimental results on real-world data that validate our theoretical model.
Co-authors: Prof. Mário Alvim, Natasha Fernandes, Prof. Annabelle McIver.
Computer Science | Topics API | Third-Party Cookies | Quantitative Information Flow | Privacy | Utility | Privacy-Utility Trade-Off | Differential Privacy

A Novel Analysis of Utility in Privacy Pipelines, Using Kronecker Products and Quantitative Information Flow — November 2023
2023 ACM SIGSAC Conference on Computer and Communications Security. **DOI**: 10.1145/3576915.3623081. **arXiv**: 2308.11110.
We combine Kronecker products, and quantitative information flow, to give a novel formal analysis for the fine-grained verification of utility in complex privacy pipelines. The combination explains a surprising anomaly in the behaviour of utility of privacy-preserving pipelines - that sometimes a reduction in privacy results also in a decrease in utility. We use the standard measure of utility for Bayesian analysis, introduced by Ghosh at al., to produce tractable and rigorous proofs of the fine-grained statistical behaviour leading to the anomaly. More generally, we offer the prospect of formal-analysis tools for utility that complement extant formal analyses of privacy. We demonstrate our results on a number of common privacy-preserving designs.
Co-authors: Prof. Mário Alvim, Natasha Fernandes, Prof. Annabelle McIver, Prof. Carroll Morgan.
Computer Science | Quantitative Information Flow | Privacy | Utility | Privacy and Utility Trade-off | Differential Privacy

Measuring Re-identification Risk — June 2023
2023 ACM SIGMOD/PODS Conference. **DOI**: 10.1145/3589294. **arXiv**: 2304.07210.
Compact user representations (such as embeddings) form the backbone of personalization services. In this work, we present a new theoretical framework to measure re-identification risk in such user representations. Our framework, based on hypothesis testing, formally bounds the probability that an attacker may be able to obtain the identity of a user from their representation. As an application, we show how our framework is general enough to model important real-world applications such as the Chrome's Topics API for interest-based advertising. We complement our theoretical bounds by showing provably good attack algorithms for re-identification that we use to estimate the re-identification risk in the Topics API. We believe this work provides a rigorous and interpretable notion of re-identification risk and a framework to measure it that can be used to inform real-world applications.
Co-authors: CJ Carey, Travis Dick, Alessandro Epasto, Adel Javanmard, Josh Karlin, Shankar Kumar, Andres Muñoz Medina, Vahab Mirrokni, Sergei Vassilvitskii, Peilin Zhong.
Computer Science | Re-identification Risk | Privacy | User Representations

Flexible and Scalable Privacy Assessment for Very Large Datasets, with an Application to Official Governmental Microdata — July 2022
22nd Privacy Enhancing Technologies Symposium (PETS 2022). **DOI**: 10.56553/popets-2022-0114. **arXiv**: 2204.13734.
We present a systematic refactoring of the conventional treatment of privacy analyses, basing it on mathematical concepts from the framework of Quantitative Information Flow (QIF). The approach we suggest brings three principal advantages: it is flexible, allowing for precise quantification and comparison of privacy risks for attacks both known and novel; it can be computationally tractable for very large, longitudinal datasets; and its results are explainable both to politicians and to the general public. We apply our approach to a very large case study: the Educational Censuses of Brazil, curated by the governmental agency INEP, which comprise over 90 attributes of approximately 50 million individuals released longitudinally every year since 2007. These datasets have only very recently (2018-2021) attracted legislation to regulate their privacy - while at the same time continuing to maintain the openness that had been sought in Brazilian society. INEP's reaction to that legislation was the genesis of our project with them. In our conclusions here we share the scientific, technical, and communication lessons we learned in the process.
Co-authors: Prof. Mário Alvim, Natasha Fernandes, Prof. Annabelle McIver, Prof. Carroll Morgan.
Computer Science | Quantitative Information Flow | Disclosure Control | Microdata | Privacy | Very Large Datasets | Longitudinal Datasets

A formal quantitative study of privacy in the publication of official educational censuses in Brazil — July 2022

Proceedings of the XXXV Thesis and Dissertation Contest (CTD 2022). DOI: 10.5753/ctd.2022.223158.

We present a summary of the work done in the dissertation *A formal quantitative study of privacy in the publication of official educational censuses in Brazil*, including its contributions and impacts so far. The dissertation presents a systematic refactoring of the conventional treatment of privacy analyses, based on mathematical concepts from the framework of Quantitative Information Flow (QIF). This brings three principal advantages: flexibility, allowing for precise quantification and comparison of privacy risks for attacks both known and novel; computational tractability for very large, longitudinal datasets; and explainable results both to politicians and to the general public. We apply our approach to a very large case study: the educational censuses in Brazil, which comprise over 90 attributes of approximately 50 million individuals released longitudinally every year since 2007.

Co-authors: Prof. Mário Alvim, Prof. Annabelle McIver.

Computer Science | Quantitative Information Flow | Disclosure Control | Microdata | Privacy | Utility | Differential Privacy

On Privacy and Accuracy in Data Releases — August 2020

31st International Conference on Concurrency Theory (CONCUR 2020). DOI: 10.4230/LIPIcs.CONCUR.2020.1.

In this paper we study the relationship between privacy and accuracy in the context of correlated datasets. We use a model of quantitative information flow to describe the the trade-off between privacy of individuals' data and and the utility of queries to that data by modelling the effectiveness of adversaries attempting to make inferences after a data release. We show that, where correlations exist in datasets, it is not possible to implement optimal noise-adding mechanisms that give the best possible accuracy or the best possible privacy in all situations. Finally we illustrate the trade-off between accuracy and privacy for local and oblivious differentially private mechanisms in terms of inference attacks on medium-scale datasets.

Co-authors: Prof. Mário Alvim, Natasha Fernandes, Prof. Annabelle McIver.

Computer Science | Privacy and Utility Trade-off | Quantitative Information Flow | Inference Attacks | Differential Privacy

</> Software

Topics API Analysis — June 2024

Zenodo. DOI: 10.5281/ZENODO.11032230.

This repository provides the experimental results of the paper *The Privacy-Utility Trade-off in the Topics API* (DOI: 10.1145/3658644.3670368; arXiv: 2406.15309).

Software | Topics API | Third-Party Cookies | Quantitative Information Flow | Privacy | Utility | Web Standards | Interest-Based Advertising

Bayes Vulnerability for Microdata (BVM) library — April 2021

Zenodo. DOI: 10.5281/zenodo.6533703.

Quantitative Information Flow assessment of vulnerability for microdata datasets using Bayes Vulnerability. This tool was used for the vulnerability assessment published in: *A formal quantitative study of privacy in the publication of official educational censuses in Brazil* (hdl:1843/38085) and *Flexible and scalable privacy assessment for very large datasets, with an application to official governmental microdata* (DOI: 10.56553/popets-2022-0114, arXiv: 2204.13734).

Software | Privacy | Utility | Formal Methods | Quantitative Information Flow | Very Large Datasets | Longitudinal Datasets

Datasets

AOL Dataset for Browsing History and Topics of Interest — June 2024

Zenodo. DOI: 10.5281/zenodo.11029571.

This record provides the datasets of the paper *The Privacy-Utility Trade-off in the Topics API* (DOI: 10.1145/3658644.3670368; arXiv: 2406.15309). The datasets generating code and the experimental results can be found in 10.5281/zenodo.11229402 (nunesgh/topics-api-analysis).

Dataset | Topics API | Third-Party Cookies | Privacy | Utility | Privacy and Utility Trade-off | Web Standards | Interest-Based Advertising

Academic Publications

A formal quantitative study of privacy in the publication of official educational censuses in Brazil — April 2021

Universidade Federal de Minas Gerais. HDL: 1843/38085.

In this thesis, we provide a thorough quantitative study of privacy risks in the release of the official Brazilian Educational Censuses provided annually by INEP, which is Brazil's governmental agency responsible for the development and maintenance of educational statistics systems. More precisely, we formally analyze privacy risks in databases released as microdata, i.e. data at each individual's record level, and protected by the technique of de-identification, i.e. the removal of direct identifying information such as the individuals' names or personal identification numbers.

Computer Science | Quantitative Information Flow | Disclosure Control | Microdata | Differential Privacy | Privacy | Utility

An Introduction to Raman Spectroscopy — November 2017

Didactic or instructional material.

Raman Spectroscopy applications are vast in Physics, Chemistry, Geology and in other areas, because it is possible to characterize different materials through their vibrational spectra. It is an efficient and non-destructive method, hence not only useful inside a laboratory, but also for some real-world problems. In this study, some of the classical Raman Spectroscopy theory shall be develop so it can be applied to a specific case on an experimental example. By the end of this study, one will have covered all the basic ideas behind the Raman Spectroscopy technique.

Supervised by Prof. Leandro Malard.

Physics | Raman Spectroscopy

A Geometric Introduction to Lie Groups — November 2016

III National Scientific Initiation and Master Program (PICME) Symposium.

Co-authors: André Nascimento, Cássio Feitosa, Cleber Barreto, Diego Carriel. Supervised by Romero Solha.

Mathematics | Lie Groups

Workshop Papers

A new Framework for Measuring Re-Identification Risk — December 2023

Workshop on Regulatable Machine Learning at the 37th Conference on Neural Information Processing Systems (RegML @ NeurIPS 2023).

Compact user representations (such as embeddings) form the backbone of personalization services. In this work, we present a new theoretical framework to measure re-identification risk in such user representations. Our framework, based on hypothesis testing, formally bounds the probability that an attacker may be able to obtain the identity of a user from their representation. As an application, we show how our framework is general enough to model important real-world applications such as the Chrome's Topics API for interest-based advertising. We complement our theoretical bounds by showing provably good attack algorithms for re-identification that we use to estimate the re-identification risk in the Topics API. We believe this work provides a rigorous and interpretable notion of re-identification risk and a framework to measure it that can be used to inform real-world applications.

Co-authors: CJ Carey, Travis Dick, Alessandro Epasto, Adel Javanmard, Josh Karlin, Shankar Kumar, Andres Muñoz Medina, Vahab Mirrokni, Sergei Vassilvitskii, Peilin Zhong.

Computer Science | Re-identification Risk | Privacy | User Representations

A Quantitative Information Flow Analysis of the Topics API — November 2023

Proceedings of the 22nd Workshop on Privacy in the Electronic Society (WPES 2023). DOI: 10.1145/3603216.3624959.

Third-party cookies have been a privacy concern since cookies were first developed in the mid 1990s, but more strict cookie policies were only introduced by Internet browser vendors in the early 2010s. More recently, due to regulatory changes, browser vendors have started to completely block third-party cookies, with both Firefox and Safari already compliant. The Topics API is being proposed by Google as an additional and less intrusive source of information for interest-based advertising (IBA), following the upcoming deprecation of third-party cookies. Initial results published by Google estimate the probability of a correct re-identification of a random individual would be below 3% while still supporting IBA. In this paper, we analyze the re-identification risk for individual Internet users introduced by the Topics API from the perspective of Quantitative Information Flow (QIF), an information- and decision-theoretic framework. Our model allows a theoretical analysis of both privacy and utility aspects of the API and their trade-off, and we show that the Topics API does have better privacy than third-party cookies. We leave the utility analyses for future work.

Co-authors: Prof. Mário Alvim, Natasha Fernandes, Prof. Annabelle McIver.

Computer Science | Topics API | Third-Party Cookies | Quantitative Information Flow | Privacy | Utility

A novel analysis of utility in privacy pipelines, using Kronecker products and quantitative information flow — September 2023

Theory and Practice of Differential Privacy (TPDP) 2023.

We combine Kronecker products, and quantitative information flow, to give a novel formal analysis for the fine-grained verification of utility in complex privacy pipelines. The combination explains a surprising anomaly in the behaviour of utility of privacy-preserving pipelines - that sometimes a reduction in privacy results also in a decrease in utility. We use the standard measure of utility for Bayesian analysis, introduced by Ghosh at al., to produce tractable and rigorous proofs of the fine-grained statistical behaviour leading to the anomaly. More generally, we offer the prospect of formal-analysis tools for utility that complement extant formal analyses of privacy. We demonstrate our results on a number of common privacy-preserving designs.

Co-authors: Prof. Mário Alvim, Natasha Fernandes, Prof. Annabelle McIver, Prof. Carroll Morgan.

Computer Science | Quantitative Information Flow | Privacy | Utility | Privacy and Utility Trade-off | Differential Privacy

Measuring Re-identification Risk — May 2023

2023 Security for the Web Workshop.

Compact user representations (such as embeddings) form the backbone of personalization services. In this work, we present a new theoretical framework to measure re-identification risk in such user representations. Our framework, based on hypothesis testing, formally bounds the probability that an attacker may be able to obtain the identity of a user from their representation. As an application, we show how our framework is general enough to model important real-world applications such as the Chrome's Topics API for interest-based advertising. We complement our theoretical bounds by showing provably good attack algorithms for re-identification that we use to estimate the re-identification risk in the Topics API. We believe this work provides a rigorous and interpretable notion of re-identification risk and a framework to measure it that can be used to inform real-world applications.

Co-authors: CJ Carey, Travis Dick, Alessandro Epasto, Adel Javanmard, Josh Karlin, Shankar Kumar, Andres Muñoz Medina, Vahab Mirrokni, Sergei Vassilvitskii, Peilin Zhong.

Computer Science | Re-identification Risk | Privacy | User Representations

Events

- 2024 ACM SIGSAC Conference on Computer and Communications Security — 2024
  CCS — *Symposium, Salt Lake City, UT, U.S.A.*
- 22nd Privacy Enhancing Technologies Symposium (YouTube) — 2022
  PETS — *Symposium. Sydney, NSW, Australia, and Online.*
- 9^{th} Summer School in Computer Science, at the Department of Computer Science (PDF) (YouTube) — 2020
  Universidade Federal de Minas Gerais — *Seminary. Belo Horizonte, MG, Brazil.*
- 8^{th} Summer School in Computer Science, at the Department of Computer Science — 2019
  Universidade Federal de Minas Gerais — *Seminary. Belo Horizonte, MG, Brazil.*
- VII Summer School in Computer Science, at the Department of Computer Science — 2018
  Universidade Federal de Minas Gerais — *Seminary. Belo Horizonte, MG, Brazil.*
- III National Scientific Initiation and Master Program (PICME) Symposium (PDF) — 2016
  Universidade Federal de Minas Gerais — *Symposium. Belo Horizonte, MG, Brazil.*
- X CBPF'S School, at the Brazilian Center for Physics Research — 2015
  Centro Brasileiro de Pesquisas Físicas — *Seminary. Rio de Janeiro, RJ, Brazil.*
- XXV School of Physics, at the Department of Physics — 2014
  Universidade Federal de Minas Gerais — *Seminary. Belo Horizonte, MG, Brazil.*

Projects

PRICE - Privacy in Educational Censuses

2020 - 2021

PRICE INSCRYPT Information Security Analyst

The new Brazilian privacy legislation legally holds entities as responsible for the quality, confidentiality, and privacy of data they keep about individuals. INEP, the National Institute of Educational Studies and Research of the Ministry of Education, publishes very detailed educational data annually. The PRICE project was a study commissioned by INEP on how to transform the data to be published in a way that the privacy of students is not violated, while maintaining its utility for statistical research.

Technical Work:

- Report 01: Report on the international panorama and the INEP context regarding methods of handling privacy control in statistical disclosure.
- Report 02: Report on the risks to privacy arising from the current form of disclosure of microdata from INEP Educational Censuses.
- Report 03: Technical report on treatment methods applicable to the dissemination microdata of INEP's Educational Censuses.
- Report 04: Technological solution and its documentation.
- Report 05: Final technical report of the pilot project.
- Report 06: Technical Implementation Report.
- Report 07: Assisted Operation Report.
- Report 08: Project closure report.

Decentralized Execution Term (TED) INEP-UFMG 8750.
Coordinator: Prof. Mário Alvim.

Computer Science | Privacy | Transparency | Python

Workshop on database anonymization techniques

November 30, 2018

Educational Statistics Directorate (DEED/INEP)
Regional Planning and Development Center (CEDEPLAR/UFMG)

The workshop aims to present INEP professionals with the state of the art of data anonymization techniques with the most recurrent use, their advantages and disadvantages, in order to support organizational decisions regarding the adoption of one or more techniques, considering the technical capacity of the teams , infrastructure, operation and any existing legal limitations.
Coordinator: Prof. Mário Alvim.

Computer Science | Privacy | Transparency

Lumus Max Optical Spectrometer

2014 - 2015

Lumus Max (in Portuguese) Room of Physics Demonstrations (in Portuguese)

Development of a high-resolution and low-cost optical spectrometer. Some parts of the hardware were designed using Dassault Systèmes' SolidWorks and the software was implemented using National Instruments' LabVIEW visual programming language.
Coordinator: Prof. Elmo Salomão.

Physics | Spectrometer | LabVIEW | SolidWorks

Electrical Paul's Trap

2014 - 2015

Electrical Paul's Trap (in Portuguese) Room of Physics Demonstrations (in Portuguese)

Development of an electrical Paul's Trap (quadrupole ion trap).
Coordinator: Prof. Elmo Salomão.

Physics