

Investigating the Perception of Brazilian Software Developers on Data Privacy in the Use of Large Language Models (LLMs)

The objective of this research is to investigate the level of knowledge and understanding that Brazilian software developers have regarding data privacy when using Large Language Models (LLMs) in their professional activities.

Research Questions (RQ):

- RQ1 - What is the level of knowledge of Brazilian software developers about basic data privacy concepts?
- RQ2 - What is the level of knowledge of Brazilian software developers about LLMs?
- RQ3 - What is the level of understanding of Brazilian software developers about the data privacy risks associated with the use of LLMs in their professional activities?
- RQ4 - What are the practices and procedures adopted by Brazilian software developers to mitigate data privacy risks when using LLMs in their professional activities?
- RQ5 - What are the challenges faced by Brazilian software developers in ensuring data privacy when using LLMs in their professional activities?
- RQ6 - What are the recommendations for ensuring data privacy when using LLMs?

PARTICIPANT PROFILE

Q01 - How old are you?

- 20 years or younger
- Between 21 and 25 years
- Between 26 and 30 years
- Between 31 and 36 years
- Between 37 and 42 years
- Between 43 and 47 years
- Between 48 and 54 years
- Between 55 and 60 years
- 61 years or older
- I prefer not to answer

Q02 - Which Brazilian state do you live in?

- AC
- AL
- AP
- AM

- BA
- CE
- DF
- ES
- GO
- MA
- MT
- MS
- MG
- PA
- PB
- PR
- PE
- PI
- RJ
- RN
- RS
- RO
- RR
- SC
- SP
- SE
- TO
- None (I live outside Brazil)

Q03 - What is your level of education?

- High school
- Graduate
- Postgraduate degree (Specialization) or MBA
- Master's degree
- Doctorate

Q04 - What is your main role currently in the software development area?

- Product Owner
- Project Manager
- Scrum Master
- Technical and/or Team Leader
- Business Analyst
- Software Requirements Analyst
- Software Architect
- Programmer and/or Developer
- Quality Analyst and/or Tester
- Designer and/or UI/UX Specialist
- Engineer and/or Data Analyst
- DevOps Engineer
- Other

Q05 - How many years of experience do you have working in software development?

- Less than 1 year
- Between 1 and 3 years
- Between 4 and 6 years
- Between 7 and 9 years
- Between 10 and 12 years
- Between 13 and 15 years
- Over 15 years

RQ1 - WHAT IS THE LEVEL OF KNOWLEDGE OF BRAZILIAN SOFTWARE DEVELOPERS ABOUT BASIC DATA PRIVACY CONCEPTS?

Q06 - What is your level of knowledge about data privacy?

- No Knowledge
- Basic Awareness (I have a general idea of what data privacy is)
- Basic Knowledge (I understand the basic principles and rights of data subjects)
- Intermediate Knowledge (I can explain the main concepts and how to apply them in common situations)
- Comprehensive Knowledge (I have a good understanding of the topic, including data privacy laws and regulations)

Q07 - Have you ever attended training or courses on data privacy?

- I have never participated
- I have participated in an introductory training
- I have participated in a basic course
- I have participated in an intermediate course
- I have participated in an advanced or specialization training

Q08 - What is your main source of information about data privacy?

- Courses
- Internet
- Books
- Magazines and articles
- Events and workshops
- Organizations and associations
- Specialized consultancies
- Contact with experts
- Legislative updates
- Other

Q09 - What is your level of knowledge about the Brazilian General Data Protection Law (LGPD)?

- No knowledge
- Basic Awareness (I have a general idea of the principles and rights)

- Basic knowledge (I understand the main points of the law and how to apply them in simple situations)
- Intermediate Knowledge (I can explain the details of the law, including business obligations and data subject rights)
- Comprehensive Knowledge (I have a command of the subject and in-depth knowledge of the legal implications of the LGPD)

Q10 - What is your level of knowledge about the principles of the Brazilian General Data Protection Law (LGPD)?

- No knowledge
- Basic Awareness (I have a general idea of the principles, but need more information)
- Basic knowledge (I understand the main principles and how they apply in simple situations)
- Intermediate Knowledge (I can explain the principles in detail and with examples, including how to identify violations of the LGPD)
- Comprehensive Knowledge (I have mastery of the subject and can explain all the principles of the LGPD in depth, including their legal implications)

Q11 - What are the main principles of the LGPD that you know? (Check all that apply)

- None
- Purpose
- Adequacy
- Need
- Open Access
- Data Quality
- Transparency
- Security
- Prevention
- Non-Discrimination
- Accountability

Q12 - Which of the following LGPD data protection concepts are you familiar with? (Check all that apply)

- Personal data
- Sensitive personal data
- Processing of personal data
- Processing of sensitive personal data
- Anonymization and/or pseudonymization of personal data
- Agents processing personal data
- Consent of the data subject
- Rights of the data subject (e.g. access, correction, deletion)
- Accountability and compensation for damages
- Penalties for non-compliance with the LGPD
- Other

Q13 - What is your level of knowledge about the differences between personal data and sensitive data?

- No knowledge

- Basic Awareness (I have a general idea, but need more information)
- Basic knowledge (I understand the main concepts and some differences)
- Intermediate Knowledge (I can explain the differences with examples and apply the knowledge in simple situations)
- Comprehensive Knowledge (I have mastered the topic and can explain the differences in detail, including legal implications)

Q14 - What is your level of knowledge about data subject rights under the LGPD?

- No knowledge
- Basic Awareness (I have a general idea of the rights, but need more information)
- Basic knowledge (I understand the main rights and how to exercise them in simple situations)
- Intermediate Knowledge (I can explain the rights in detail and with examples, including how to request access to my data)
- Comprehensive Knowledge (I have mastered the subject and can explain all the rights of data subjects, including how to appeal in case of violations)

Q15 - How would you ensure data privacy in a software system?

Open question

RQ2 - WHAT IS THE LEVEL OF KNOWLEDGE OF BRAZILIAN SOFTWARE DEVELOPERS ABOUT LLMS?

Q16 - How would you rate your knowledge of LLMs?

- No knowledge
- Basic (I have a general idea)
- Intermediate (I understand the basic operation)
- Advanced (I have knowledge of applications and implications)
- Expert (I have mastery of the subject and in-depth technical knowledge)

Q17 - Which LLMs have you used or are familiar with? (Check all that apply)

- GPT by OpenAI
- Gemini by Google
- Orca or Phi-1 by Microsoft
- Llama by Meta
- Falcon by TII UAE
- Mixtral 7B by Mistral AI
- Claude by Anthropic
- StableLM 2 by Stability AI
- Coral by Cohere
- XGen-7B by Salesforce
- Sabiá by Maritaca AI
- BERTimbau by Neuralmind
- Cabrita
- BERTaú by Itaú
- Bloom by BigScience
- Grok by xAI

- Other

Q18 - Have you ever attended training or courses on LLMs?

- I have never participated
- I have participated in an introductory training
- I have participated in a basic course
- I have participated in an intermediate course
- I have participated in an advanced or specialization training

Q19 - How do you use LLMs in your professional activities?

Open question

Q20 - In your opinion, do you have a good understanding of the capabilities and limitations of LLMs?

- No knowledge
- Basic Awareness (I have a general idea of some of the basic capabilities and limitations of LLMs)
- Basic knowledge (I understand the main capabilities and limitations of LLMs, but I still have questions about some aspects)
- Intermediate Knowledge (I have a good understanding of the capabilities and limitations of LLMs, including their applications and challenges)
- Comprehensive Knowledge (I have mastery of the topic, with in-depth knowledge of the capabilities, limitations and implications of LLMs)

**Q21 - Which of the following LLM applications do you know or have used?
(Check all that apply)**

- Text generation
- Machine translation
- Question answering
- Text summaries
- Text analysis
- Sentiment analysis
- Chatbots and virtual assistants
- Content creation
- Coding and programming
- Research and data analysis
- Other

RQ3 - WHAT IS THE LEVEL OF UNDERSTANDING OF BRAZILIAN SOFTWARE DEVELOPERS ABOUT THE DATA PRIVACY RISKS ASSOCIATED WITH THE USE OF LLMS IN THEIR PROFESSIONAL ACTIVITIES?

Q22 - How aware are you of the privacy risks associated with the use of LLMS?

- No knowledge
- Basic Awareness (I have a general idea of the risks)
- Basic knowledge (I understand the main risks and their impacts)
- Intermediate Knowledge (I can explain the risks in detail and with examples)
- Comprehensive Knowledge (I have mastered the topic and can discuss the nuances and implications)

Q23 - Which of the following privacy risks do you associate with the use of LLMS? (Check all that apply)

- Leakage of sensitive data
- Misuse of personal information
- Exposure to cyberattacks
- Unauthorized disclosure of personal information
- Lack of control over data used
- Difficulty in ensuring compliance with privacy regulations
- Risks of bias and discrimination
- Inadequate data storage
- Failures in data anonymization
- Unauthorized access to data
- Failures in implementing privacy policies
- Misuse of data by third parties
- Lack of transparency about data use
- Lack of control over shared data
- Collection and analysis of personal data without consent
- Other

Q24 - What are the main privacy risks you associate with the use of LLMS?

Open question

Q25 - In your opinion, what is the level of privacy risk associated with the use of LLMS in your projects?

- No risk (LLMs pose no risk to data privacy)
- Low risk (LLMs pose minimal privacy risks if used responsibly)
- Moderate risk (LLMs pose some privacy risks, but can be managed with appropriate measures)
- High risk (LLMs pose significant privacy risks and robust control measures are required)
- Very high risk (LLMs pose a significant risk to data privacy and require stringent mitigation measures)

Q26 - What is your opinion on training LLMs without compromising data privacy?

- Impossible (Training LLMs will always compromise data privacy to some degree)
- Unlikely (It is unlikely that LLMs can be trained without compromising data privacy, but there are some measures that can mitigate this risk)
- Neutral (I am not sure if LLMs can be trained without compromising data privacy)
- Possible (It is possible to train LLMs without compromising data privacy, but this requires rigorous security techniques and measures)
- Highly Likely (I believe that LLMs can be trained safely and responsibly without compromising data privacy)

Q27 - What types of personal data do you consider most sensitive to being processed by Large Language Models (LLMs)? (Check all that apply)

- Personally identifiable information (information that allows you to identify yourself)
- Demographic data
- Health and well-being data
- Financial data
- Location data
- Behaviour and browsing data
- Minority group data
- Belief and opinion data
- Children and adolescent data
- Other

Q28 - How do you assess the likelihood of data breaches occurring related to the use of LLMs?

- Very Low (I believe that it is extremely unlikely that data breaches related to the use of LLMs will occur)
- Low (I believe that it is unlikely that data breaches related to the use of LLMs will occur)
- Moderate (I believe that the likelihood of data breaches related to the use of LLMs is moderate)
- High (I believe that it is likely that data breaches related to the use of LLMs will occur)
- Very High (I believe that it is very likely that data breaches related to the use of LLMs will occur)

Q29 - How concerned are you about the impact of data breaches related to the use of LLMs?

- Not at all concerned (I am not concerned about the impact of data breaches on LLMs)
- Somewhat concerned (I am not very concerned about the impact of data breaches on LLMs)
- Moderately concerned (I have some concerns about the impact of data breaches on LLMs, but do not consider it a serious risk)
- Concerned (I am quite concerned about the potential impact of data breaches on LLMs)
- Very concerned (I am extremely concerned about the potential consequences of data breaches on LLMs)

Q30 - What steps do you take to mitigate privacy risks when using LLMs?

Open question

RQ4 - WHAT ARE THE PRACTICES AND PROCEDURES ADOPTED BY BRAZILIAN SOFTWARE DEVELOPERS TO MITIGATE DATA PRIVACY RISKS WHEN USING LLMS IN THEIR PROFESSIONAL ACTIVITIES?

Q31 - To protect your privacy when using LLMs, do you adopt any specific practices? (Check all that apply)

- I limit the data I share with LLMs
- I use LLMs from trusted sources
- I check the privacy settings of LLMs
- I monitor how my data is being used by LLMs
- I implement additional security measures, such as data anonymization
- I do not use LLMs for sensitive tasks
- Other

Q32 - How do you handle sensitive data when using LLMs?

- I do not use LLMs with sensitive data
- I use LLMs with sensitive data only in exceptional cases
- I use LLMs with sensitive data, taking basic precautionary measures, such as anonymization
- I use LLMs in secure and isolated environments, with restricted access and
- I use LLMs with sensitive data only after careful assessment of risks and benefits.

Q33 - Before using your data with LLMs, do you take any measures to protect it, such as anonymization or pseudonymization?

- I never anonymize or pseudonymize my data
- I anonymize or pseudonymize some of my data
- I anonymize or pseudonymize most of my data
- I anonymize or pseudonymize all of my data
- I implement complex and customized anonymization/pseudonymization measures

Q34 - What are the main challenges you face when implementing privacy practices when using LLMs? (Check all that apply)

- Lack of clarity in privacy guidelines (laws and regulations)
- Challenges in compliance with privacy laws and regulations
- Ethical and legal uncertainties
- Challenges in data anonymization
- Difficulty in identifying and classifying sensitive data
- Limitations of LLM technology to handle sensitive data securely and ethically
- Lack of transparency of LLMs
- Lack of tools and resources in implementing privacy practices with LLMs
- Concerns about security and inappropriate access of sensitive data when using LLMs
- Lack of knowledge and expertise in privacy and LLMs
- Difficulty in measuring the impact of privacy practices
- Lack of awareness and training
- Lack of privacy culture in the organization
- Other

RQ5 - WHAT ARE THE CHALLENGES FACED BY BRAZILIAN SOFTWARE DEVELOPERS IN ENSURING DATA PRIVACY WHEN USING LLMS IN THEIR PROFESSIONAL ACTIVITIES?

Q35 - Which of the following challenges do you find most difficult when ensuring data privacy when using LLMs in your professional activities? (Check all that apply)

- Identify and classify sensitive data
- Adopt anonymization and pseudonymization measures
- Minimize data collection and retention
- Perform privacy impact assessments
- Achieve transparency on LLM use
- Comply with data privacy laws and regulations
- Ensure accountability for LLM use
- Other

Q36 - Do you think current legislation is adequate to protect data privacy when using LLMs?

- Strongly Disagree (I believe that the current legislation is inadequate to protect data privacy when using LLMs and needs a complete overhaul)
- Somewhat Disagree (I believe that the current legislation is not sufficient to protect data privacy when using LLMs and needs to be strengthened)
- Neutral (I am not sure whether the current legislation is adequate to protect data privacy when using LLMs)
- Somewhat Agree (I believe that the current legislation provides a good level of protection for data privacy when using LLMs, but some improvements could be made)
- Strongly Agree (I believe that the current legislation provides comprehensive and effective protection for data privacy when using LLMs)

Q37 - Are you having trouble finding information or resources about data privacy when using LLMs?

- Strongly Disagree (I have no difficulty finding information and resources on how to ensure data privacy when using LLMs)
- Somewhat Disagree (I have a little difficulty finding information and resources on how to ensure data privacy when using LLMs)
- Neutral (I am not sure if I have difficulty finding information and resources on how to ensure data privacy when using LLMs)
- Somewhat Agree (I have some difficulty finding information and resources on how to ensure data privacy when using LLMs)
- Strongly Agree (I have a lot of difficulty finding information and resources on how to ensure data privacy when using LLMs)

Q38 - How does the lack of specific knowledge about data privacy affect your work with LLMs? (Check all that apply)

- Difficulty identifying and classifying sensitive data.
- Uncertainty about best practices for data protection
- Difficulty using appropriate security measures
- Risk of not complying with data privacy laws and regulations

- Difficulty assessing privacy impacts
- Uncertainty about liability for use of LLMs
- Risk of data breaches
- Uncertainty about compliance with laws and regulations
- Limited ability to use LLMs safely and responsibly
- Other

RQ6 - WHAT ARE THE RECOMMENDATIONS FOR ENSURING DATA PRIVACY WHEN USING LLMS?

Q39 - What recommendations would you give to other developers on how to protect data privacy when using LLMs?

Open Question

Q40 - In your opinion, are more trainings and courses on data privacy and LLMs needed?

- Not necessary.
- Slightly necessary
- Neutral
- Necessary
- Very necessary

Q41 - In your opinion, how important is it to have an organizational culture that addresses data privacy when using LLMs?

- Not important
- Slightly important
- Neutral
- Important:
- Very important

Q42 - What kind of support or additional resources do you think would be helpful to improve data privacy protection when using LLMs?(Check all that apply)

- Clear guidelines and best practices
- Privacy tools and technologies
- Privacy training and capacity building
- Public awareness and education campaigns
- Legal and regulatory guidance
- Privacy assessment solutions
- Privacy research and development for LLMs
- Collaboration and knowledge sharing forums
- Other