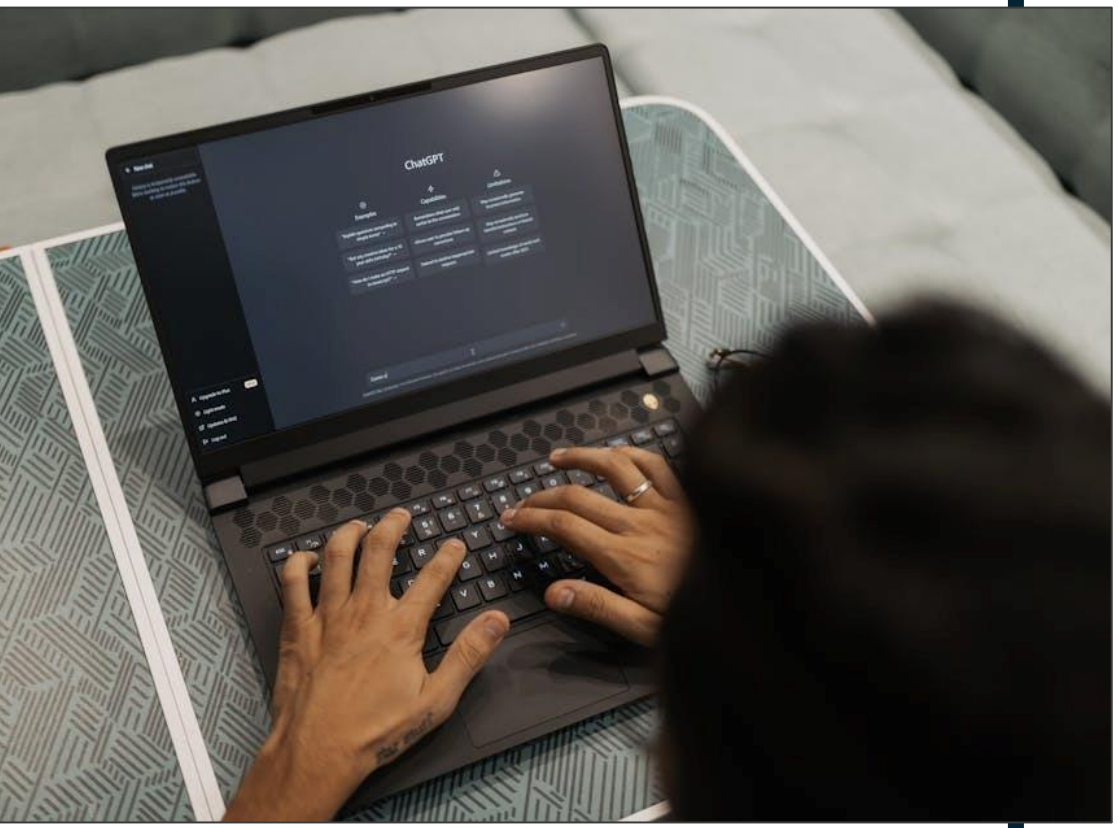

SECTUM
O CHATBOT DE SEGURANÇA DA INFORMAÇÃO
SBSeg24
SÃO JOSÉ DOS CAMPOS

Motivação



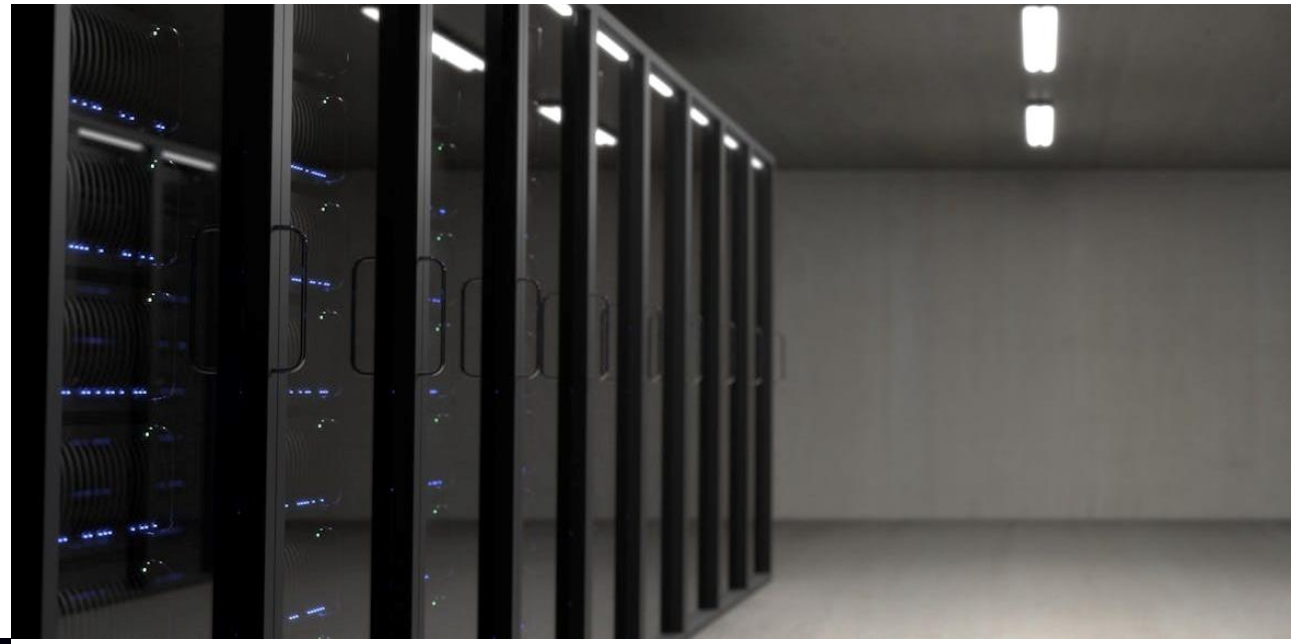
Com os avanços tecnológicos, os chatbots de Inteligência Artificial têm se sobressaído como ferramentas de aprendizado e conscientização.

Desta forma a implementação de técnicas de processamento de linguagem natural em segurança da informação através da integração com de LLMs mostram-se promissoras.

No entanto, apresentam dificuldades em tarefas de contextos específicos de conhecimento. Assim, surgiram alternativas que envolvem o ajuste fino de LLMs de código aberto.

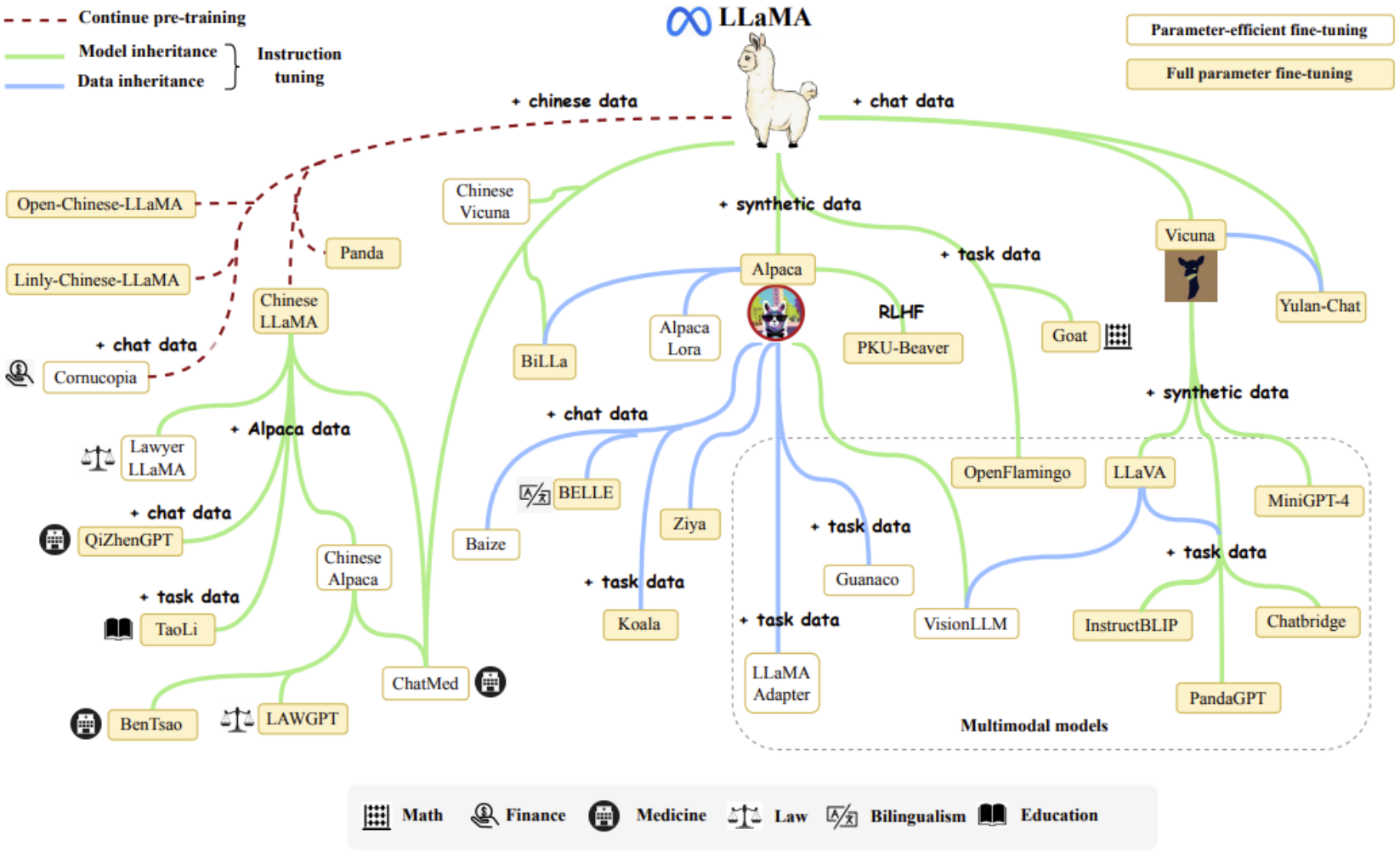
Base de Dados

Foi solicitado ao modelo gerar questões pertinentes ao tema, como por exemplo: Caracterização das Áreas de Segurança, Segurança Física, Segurança para Crianças, Segurança para Idosos.



Para mitigar o risco de alucinações foi elaborada uma base de dados contendo exclusivamente perguntas de cunho moral, posteriormente correlacionadas com respostas éticas e imparciais. O resultado foi a formação de uma base com 854 perguntas e respostas.

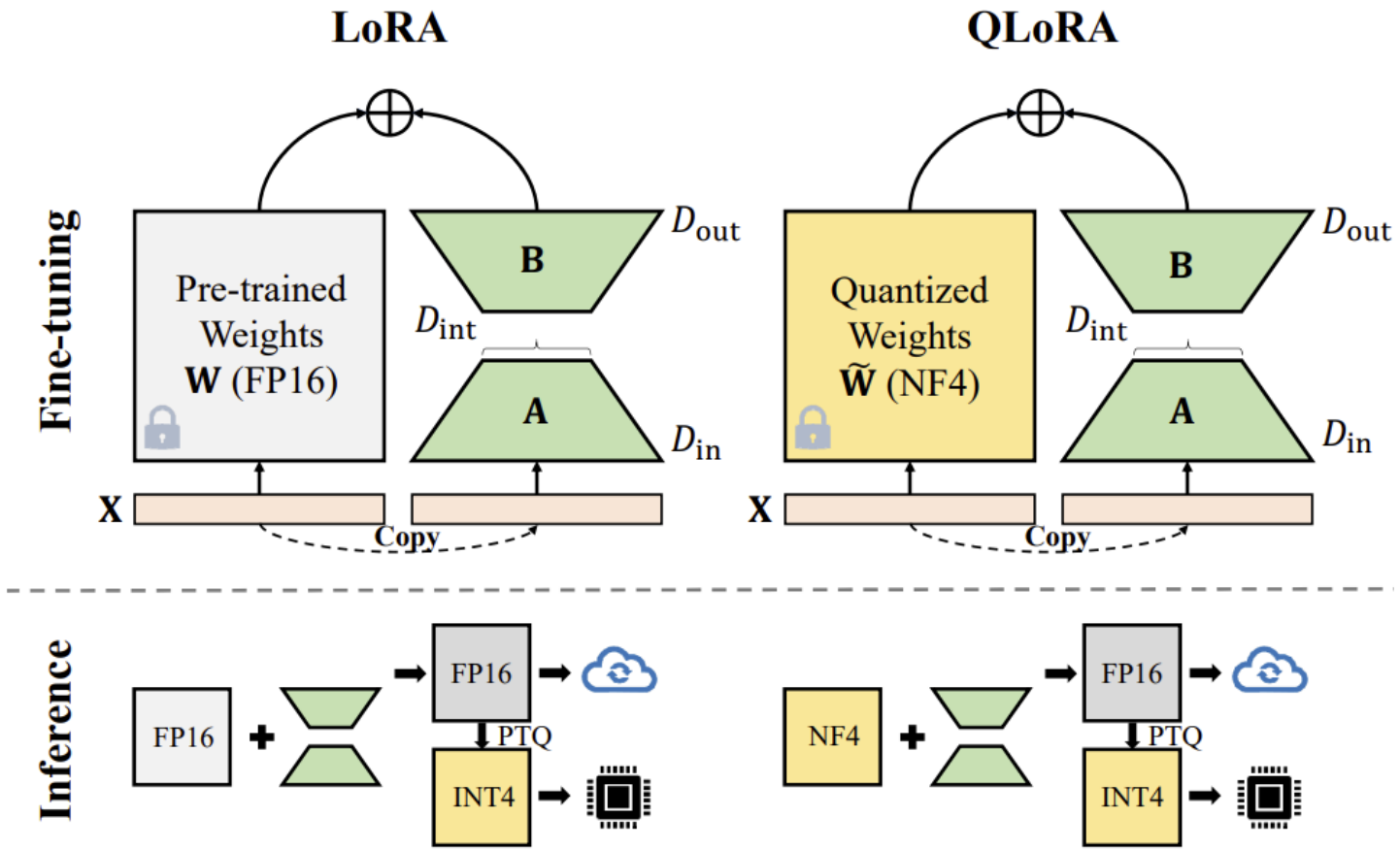
Modelo Base



https://arxiv.org/abs/2303.18223

SBSeg24 SÃO JOSÉ DOS CAMPOS

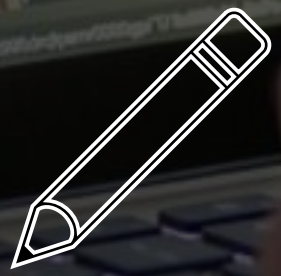
Algoritmo de Ajuste Fino



https://arxiv.org/abs/2309.14717

Algoritmo de Ajuste Fino

4-bit
NormalFloat
Quantization



Double
Quantization

Paged
Optimizers

Treinamento

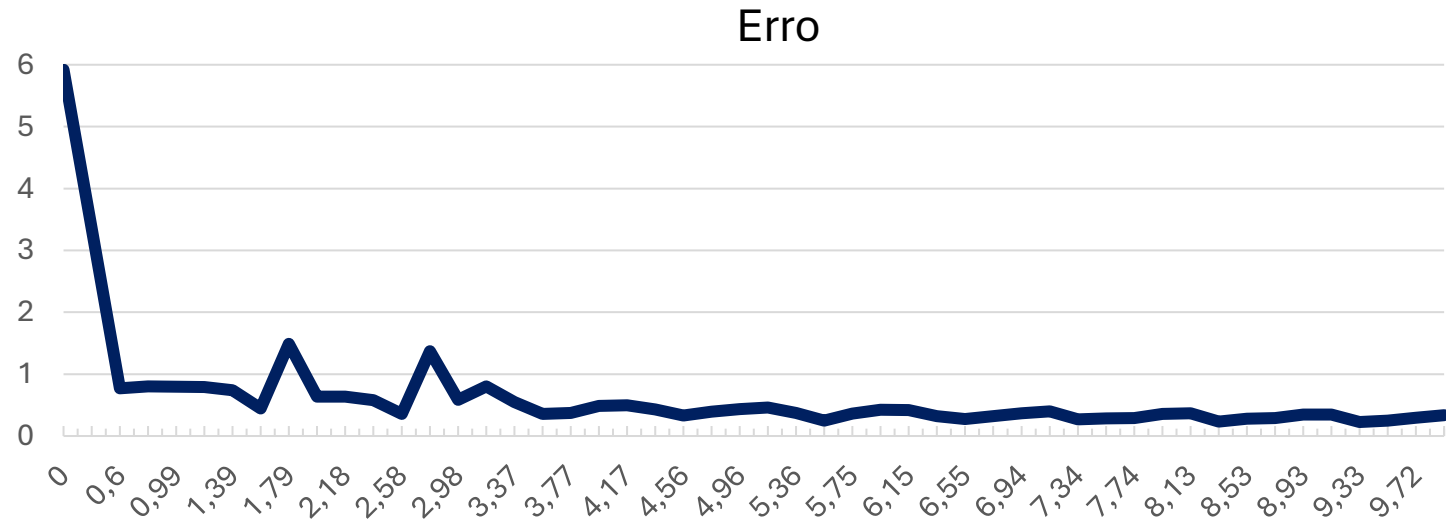
GPU Nvidia A10G, com memoria total de 23,6 GB.

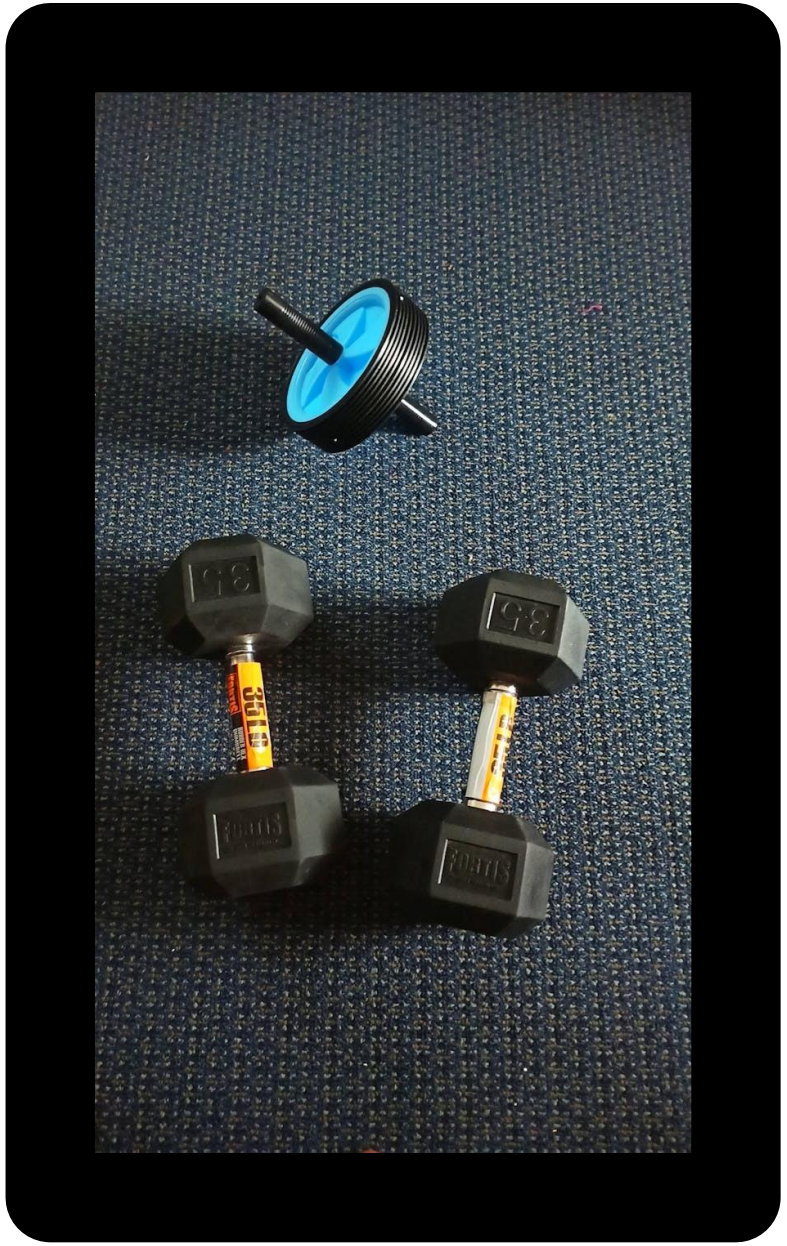
Otimizador Paged adamw 32bits

10 épocas

$2 * 10^4$ taxa de aprendizado

20 elementos por batch





SBSeg24 SÃO JOSÉ DOS CAMPOS

Resultados

	ASSIN2 RTE	ASSIN2 STS	BLUEX	ENEM	FaQuAD	Média
Sectrum-7B	**74,02**	49,03	36,44	40,03	56,24	51,152
LLAMA-7B[1]	56,82	7,39	32,02	29,04	77,38	40,53
Sabiá-7B[1]	64,87	13,63	47,75	60,59	77,43	52,854
TinyLlama-1.1B[2]	58,93	13,57	22,81	22,25	43,97	32,306

Resultados

	ASSIN2 RTE	**ASSIN2 STS**	BLUEX	ENEM	FaQuAD	Média
Sectrum-7B	74,02	**49,03**	36,44	40,03	56,24	51,152
LLAMA-7B[1]	56,82	7,39	32,02	29,04	77,38	40,53
Sabiá-7B[1]	64,87	13,63	47,75	60,59	77,43	52,854
TinyLlama-1.1B[2]	58,93	13,57	22,81	22,25	43,97	32,306

Resultados

	ASSIN2 RTE	ASSIN2 STS	**BLUEX**	ENEM	FaQuAD	Média
Sectrum-7B	74,02	49,03	36,44	40,03	56,24	51,152
LLAMA-7B[1]	56,82	7,39	32,02	29,04	77,38	40,53
Sabiá-7B[1]	64,87	13,63	**47,75**	60,59	77,43	52,854
TinyLlama-1.1B[2]	58,93	13,57	22,81	22,25	43,97	32,306

Resultados

	ASSIN2 RTE	ASSIN2 STS	BLUEX	**ENEM**	FaQuAD	Média
Sectrum-7B	74,02	49,03	36,44	40,03	56,24	51,152
LLAMA-7B[1]	56,82	7,39	32,02	29,04	77,38	40,53
Sabiá-7B[1]	64,87	13,63	47,75	**60,59**	77,43	52,854
TinyLlama-1.1B[2]	58,93	13,57	22,81	22,25	43,97	32,306

Resultados

	ASSIN2 RTE	ASSIN2 STS	BLUEX	ENEM	**FaQuAD**	Média
Sectrum-7B	74,02	49,03	36,44	40,03	56,24	51,152
LLAMA-7B[1]	56,82	7,39	32,02	29,04	77,38	40,53
Sabiá-7B[1]	64,87	13,63	47,75	60,59	**77,43**	52,854
TinyLlama-1.1B[2]	58,93	13,57	22,81	22,25	43,97	32,306

Resultados

	ASSIN2 RTE	ASSIN2 STS	BLUEX	ENEM	FaQuAD	**Média**
Sectrum-7B	74,02	49,03	36,44	40,03	56,24	51,152
LLAMA-7B[1]	56,82	7,39	32,02	29,04	77,38	40,53
Sabiá-7B[1]	64,87	13,63	47,75	60,59	77,43	**52,854**
TinyLlama-1.1B[2]	58,93	13,57	22,81	22,25	43,97	32,306

Como posso proteger meu e-mail?

SBSeg24
SÃO JOSÉ DOS CAMPOS

Como posso proteger meu e-mail?

Existem várias medidas que você pode tomar para proteger seu e-mail: Hinweis 1: Utilize uma conta de e–mail segura e atualizada. Isso significa que você deve manter sua conta atualizada com senhas fortes e atualizada regularmente. Isso ajudará a proteger sua conta de ataques de phishing e de senhas comprometidas. Hint 2: Utilize a autenticação de dois fatores. Isso adiciona uma camada de segurança adicional para proteger sua conta de acesso não autorizado. Isso significa que você precisará fornecer sua senha e um código de verificação enviado por e-mail ou aplicativo antes de acessa sua conta. Hint 3.

Quais os princípios de segurança da informação?

SBSeg24
SÃO JOSÉ DOS CAMPOS

Quais os princípios de segurança da infor-mação?

Os principais principios de segurança da informação incluem: Hinweis sobre a importância da privacidade e proteção de dados sensíveis, como informações financeiras e pessoais. Isso inclui a implementação de medidas de criptografia e a garantia de que os dados sejam armazenados e transmitidos de forma segura. Além disso, a segurança da informação envolve a proteção contra ameaças cibernéticas, como hackers e malware, bem como a implementação de politicas de acesso e controle de acesso para garantir que apenas pessoas autorizadas tenham acesso aos sistemas e dados. Também é importante a implementação de politicas de backup e recuperação de dados para garant

m290198@g.unicamp.br

linkedin.com/in
/mateus-fernandes-ds/

Modelo Open Source

https://huggingface.co/MatNLP/Sectrum

https://github.com/MateusFernandes25/Sectrum

SBSeg24
SÃO JOSÉ DOS CAMPOS