Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Análise de ética computacional no projeto de uma sociedade artificial multi-agente para ecossistema educacional

Wanderlan Alves de Jesus Brito

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia de Computação

Orientadora
Prof.a Germana Menezes da Nobrega

Brasília
2023

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Análise de ética computacional no projeto de uma sociedade artificial multi-agente para ecossistema educacional

Wanderlan Alves de Jesus Brito

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia de Computação

Prof.a Germana Menezes da Nobrega (Orientadora)
UnB/CIC

Prof. Dr. André von Borries Lopes
UnB/ENM

Prof.a Dr.a Edna Dias Canedo
UnB/CIC

Prof. João Luiz Azevedo de Carvalho
Coordenador do Curso de Engenharia de Computação

Brasília, 4 de Dezembro de 2023

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

Alves de Jesus Brito, Wanderlan
AA474a Análise de ética computacional no projeto de uma sociedade artificial multi-agente para ecossistema educacional / Wanderlan Alves de Jesus Brito; orientador Germana Menezes da Nobrega. -- Brasília, 2023.
45 p.

Monografia (Graduação - Engenharia de Computação) -- Universidade de Brasília, 2023.

1. Inteligência artificial. 2. Princípios éticos. 3. Sistemas multi-agentes. 4. SmartUnB.ECOS. 5. CICFriend. I. Menezes da Nobrega, Germana, orient. II. Título.

Dedicatória

Dedico este trabalho à toda comunidade acadêmica da Universidade de Brasília, que estão em constante esforço para aperfeiçoar a universidade e torná-la um ambiente cada vez melhor. Gostaria de dedicar também aos meus pais, que forneceram todas as condições necessárias para que eu pudesse realizar meus estudos de forma confortável. E por fim aos meus amigos, que me apoiaram e me auxiliaram ao longo da minha graduação.

Agradecimentos

Agradeço a Universidade de Brasília por proporcionar toda infraestrutura necessária para que fosse possível a realização deste trabalho, assim como a realização de toda a minha graduação.

Agradeço também a professora Germana Menezes, minha orientadora, por me auxiliar, tanto na decisão do tema realizado, quanto na pesquisa necessária para a execução deste trabalho.

Por fim, agradeço a minha amiga Lúcia, que muito me incentivou e me auxiliou ao longo do desenvolvimento deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

Com o crescente desenvolvimento das tecnologias de inteligência artificial tem crescido também as discussões em torno das possíveis violações éticas que dela podem advir, já que parte do funcionamento desse tipo de tecnologia consiste em coletar e manipular dados de seus usuários. Dados esses que muitas vezes acabam sendo obtidos de formas pouco transparentes ou até mesmo sem o consentimento dos mesmos. Em relação ao ambiente educacional, foco deste estudo, quando se utiliza tecnologias de inteligência artificial os dados em questão são as informações pessoais de alunos e professores que utilizem o sistema.

Diante dessa preocupação e dado o atual cenário do projeto SmartUnB.ECOS, este trabalho propõe a descrição de um modelo AGR, para ser utilizado de referência no desenvolvimento do SmartUnB.ECOS, e faz uma análise ética encima do modelo proposto, com o objetivo de se identificar os riscos e os benefícios da utilização de agentes autônomos no ambiente educacional em questão.

Para isso, o trabalho correlaciona algumas funcionalidades do modelo AGR proposto com as principais etapas de manipulação de dados e com alguns princípios éticos, considerados essenciais para a utilização benéfica das inteligências artificiais. Dessa forma o trabalho evidencia alguns dos riscos mais evidentes da utilização do modelo AGR no projeto SmartUnB.ECOS e levanta a atenção para os princípios éticos que não podem ser ignorados ao longo do desenvolvimento do projeto como um todo.

Palavras-chave: ética, inteligência artificial, princípios éticos, sistemas multi-agentes, SmartUnB, CICFriend

Abstract

With the growing development of artificial intelligence technologies, discussions around potential ethical violations have also increased. This is because a significant part of the operation of such technology involves collecting and manipulating data from its users. Often, these data are acquired in ways that lack transparency or even without the users' consent.

In the context of the educational environment, the focus of this study, when artificial intelligence technologies are employed, the data in question are the personal information of students and teachers who use the system.

Given this concern and considering the current scenario of the SmartUnB.ECOS project, this work proposes the description of an AGR model, to be used as a reference in the development of SmartUnB.ECOS, and conducts an ethical analysis of the proposed model with the aim of identifying the risks and benefits of using autonomous agents in the specific educational environment.

To achieve this, the study correlates specific functionalities of the proposed AGR model with the primary stages of data manipulation and with certain ethical principles deemed essential for the beneficial use of artificial intelligence. In doing so, the work highlights some of the most apparent risks associated with the utilization of the AGR model in the SmartUnB.ECOS project and draws attention to ethical principles that must not be overlooked throughout the development of the project as a whole.

Keywords: ethic, artificial intelligence, ethical principles, multi-agent systems, SmartUnB, CICFriend

Sumário

Lista de Figuras

Capítulo 1

Introdução

1.1 Contextualização

É evidente que quanto mais se avança com o desenvolvimento de novas tecnologias, mais comum se torna a utilização da inteligência artificial (IA) para auxiliar diversos setores da vida humana. As aplicações de IA vão desde áreas específicas, como a medicina [5], até implementações muito abrangentes e complexas, como sistemas de segurança em cidades [6]. Mas, na medida que o uso de algoritmos de IA cresce, aumentam também os problemas e falhas geradas por eles [7].

Um dos problemas mais frequentes encontrados em sistemas IA são os que envolvem questões relativas à segurança e privacidade dos dados dos usuários [8], como por exemplo, as constantes polêmicas envolvendo vazamentos de dados de reconhecimento facial por meio do uso de aplicativos [9], que levam a uma das principais discussões referentes ao desenvolvimento de novas IA, que é a que se relaciona ao papel da ética no uso deste tipo de tecnologia.

Dado esse cenário, a comunidade de IA no contexto da educação tem alertado quanto a necessidade de se incorporar análises éticas em seus projetos [10, 11, 12, 13]. Ética pode ser compreendida como um conjunto de regras e preceitos relativos à moral de um indivíduo, ou de uma sociedade. No contexto da computação, o conceito de ética está relacionado com a forma como a tecnologia afeta o dia a dia das pessoas e com quais justificativas essas tecnologias são utilizadas [14]. Nesse contexto, grande parte da literatura com foco na área da ética em IA apresenta discussões teóricas à respeito dos problemas mais comuns que surgem com o uso da tecnologia, e quais seriam os princípios que deveriam nortear os sistemas e seus desenvolvedores, para que esses problemas pudessem ser evitados [9].

O presente estudo tem, como um de seus objetivos, entender quais devem ser os principais princípios a serem considerados durante o desenvolvimento de uma IA em ambientes educacionais, mais especificamente dentro de um ecossistema digital da Universidade de

Brasília. Trabalhos anteriores [1, 15, 16] se preocuparam em coletar e analisar dados à respeito de diretrizes existentes no campo da ética computacional, evidenciando alguns princípios éticos que foram identificados como sendo os mais necessários e relevantes. Esses princípios serão melhor explicados ao longo dos próximos capítulos.

1.2 Justificativa

Para que os usuários tenham confiança e conforto ao utilizar serviços que funcionam à base de algoritmos inteligentes, é necessário que todo o desenvolvimento e funcionamento desses algoritmos seja transparente, e trabalhem com sistemas rígidos de segurança e privacidade.

Um dos subcampos de IA é a inteligência artificial distribuída (IAD), que consiste em tentar transformar um problema complexo, normalmente envolvendo grandes quantidades de dados, em problemas menores e mais simples. Uma das formas de se realizar essa transformação é com o emprego de sistemas multi-agentes (SMA), que consistem no uso de agentes autônomos, capazes de interagir com o ambiente e entre si, aprendendo e tomando suas próprias decisões [17].

Para que um SMA seja eficiente e funcione da melhor forma possível os agentes precisam ter acesso a diversos dados dos usuários do sistema. No caso do sistema educacional universitário, foco deste trabalho, os agentes devem conhecer os dados pessoais dos alunos, tais como: nome, matrícula e idade, além de outras informações relacionadas à graduação, tais como: período atual, disciplinas em curso, colegas em situações semelhantes, rendimento, entre outros aspectos que serão oportunamente abordados no decorrer do estudo.

Destarte, surge uma questão norteadora da pesquisa: de que forma um sistema como um SMA pode garantir que seus usuários tenham suas informações e sua privacidade preservadas, seguindo todos os onze princípios éticos necessários? Esta questão busca auxiliar desenvolvedores e outros pesquisadores a se atentarem para as questões éticas envolvidas na construção de sistemas de IA, que têm por finalidade interagir com pessoas.

Este trabalho se situa dentro do contexto do projeto SmartUnB.ECOS [4], um ecossistema educacional digital para socialização e aprendizagem no âmbito da Universidade de Brasília (UNB), usando o ecossistema como objeto de estudo para o desenvolvimento e análise ética de um SMA.

O ecossistema é dividido em diversos elementos e tem sido desenvolvido e aperfeiçoado pelos próprios integrantes do departamento de computação (CIC) da UnB. Desta forma espera-se que este trabalho, sendo pioneiro em incorporar uma análise ética ao design de

uma SMA do ecossistema, sirva como base teórica para trabalhos futuros e ajude a validar a matriz de riscos e benefícios feita por Rangel [2].

1.3 Problema de pesquisa

Por serem recentes os problemas e as polêmicas éticas relacionadas às tecnologias de inteligência artificial, ainda não se estabeleceram técnicas eficazes para se lidar com essa questão específica.

Recentemente a equipe de pesquisa do projeto SmartUnB.ECOS colocou a rede social CICFriend [18, 19] à disposição do público. A rede pretende utilizar a tecnologia SMA para desenvolver assistentes virtuais para os alunos, porém até então poucos trabalhos foram desenvolvidos com foco em analisar as questões éticas envolvidas com o uso de assistentes virtuais e a problemática relacionada a privacidade e segurança dos alunos.

Por opção metodológica, no decorrer do presente estudo, para se referir ao ambiente virtual de multi-agentes, será usado o nome simplificado de *Kirilândia* (nome dado ao ambiente virtual de multi-agentes pelos próprios integrantes do projeto).

1.4 Objetivos

1.4.1 Objetivo geral

O trabalho tem como objetivo analisar as questões de ética envolvidas na utilização de sistemas IA em ambiente educacional, mais especificamente na *Kirilândia*, em busca de se traçar paralelos entre os onze princípios éticos mencionados em 1.1 e algumas das funcionalidades da *Kirilândia*, identificando entre eles, aqueles mais relevantes e que devem ser considerados ao longo do desenvolvimento do sistema.

1.4.2 Objetivos específicos

Design organizacional de um SMA

Realizar a descrição básica dos agentes do ambiente da *Kirilândia* com base no modelo *agents groups and roles* (AGR) em livre tradução do inglês para agentes, grupos e papéis [20].

Análise pela matriz de riscos e benefícios

Analisar os riscos e benefícios através da descrição básica, usando como base a matriz de riscos e benefícios do estudo de Rangel [2].

Análise pelos onze princípios éticos

Analisar o ambiente da *Kirilândia* tendo como base os onze princípios éticos elencados pelo estudo [1].

1.5 Metodologia

1.5.1 Suposições/Hipóteses

É suposto que seria possível por meio do método AGR descrever de forma genérica o ambiente dos agentes autônomos da *Kirilândia*, e a partir desta descrição, realizar a análise ética considerando os onze princípios de [1] e a matriz de riscos e benefícios do estudo de Rangel.

1.5.2 Classificação da Pesquisa

O presente trabalho é um estudo transversal, qualitativo, interpretativo que tem como cenário o ambiente do projeto SmartUnB.ECOS da UnB e consistirá da aplicação de princípios gerais da ética na computação ao cenário específico estudado.

1.5.3 Delimitação

O trabalho se propõe a fazer uma análise apenas quanto ao que diz respeito às questões éticas da *Kirilândia* e consequentemente de sistemas multi-agentes em ambientes educacionais como um todo. Dessa forma não há qualquer intenção de se desenvolver o design final que será utilizado no sistema.

Também não faz parte do escopo do trabalho discutir os conceitos filosóficos quanto à ética dos princípios que aqui serão usados como parâmetro. O objetivo aqui é puramente criar relações entre os princípios em questão e a *Kirilândia*.

1.6 Organização do documento

Este trabalho está organizado em 5 capítulos, incluindo este capítulo de introdução. O capítulo 2, destina-se a apresentação de estudos desenvolvidos nas últimas décadas e que abordam temas relacionados ao uso da tecnologia em ambientes de aprendizagem.

No capítulo 3 está a revisão do material literário utilizado como base para este trabalho, assim como as contextualizações dos conceitos aqui utilizados.

O capítulo 4 traz a proposta do trabalho em si com a descrição de um SMA para a *Kirilândia* seguida pela análise ética do que foi descrito.

Por fim o capítulo 5 traz a conclusão do trabalho, com os objetivos alcançados e a necessidade de trabalhos futuros.

Capítulo 2

Trabalhos relacionados a assistentes inteligentes em ambientes educacionais

Este capítulo está dividido em quatro seções, a seção 2.1 traz aspectos gerais e abrangentes em relação ao conceito de agentes autônomos. A seção 2.2 aborda a utilização de agentes pedagógicos como forma de enriquecer e motivar o aprendizado. A seção 2.3 é dedicada a falar sobre a utilização de SMAs em ambientes educacionais. Por fim, a seção 2.4 apresenta as considerações finais do capítulo.

2.1 Generalidades sobre agentes autônomos

Segundo o livro [20], de Wooldridge, não há ainda um consenso geral quanto a definição de um agente autônomo. Parte da dificuldade para se estabelecer um conceito universal está no fato de que diferentes contextos podem exigir diferentes características para os agentes. Porém como o próprio nome indica, a autonomia é uma característica fundamental para defini-los.

Para alguns tipos de aplicações é essencial que os agentes possuam a capacidade de aprender a partir de suas experiências, para outras aplicações, no entanto, a habilidade de aprendizado dos agentes pode ser uma habilidade sem muita importância ou até mesmo indesejada.

Destarte, uma definição suficientemente satisfatória para o presente estudo é a dada pelo mesmo autor no livro [21]: "Um *agente* é um sistema computacional que está *situado* em algum *ambiente*, sendo capaz de executar *ações autônomas* nesse ambiente para alcançar seus respectivos objetivos."(Tradução livre).

2.2 Agentes pedagógicos

Alguns estudos das últimas décadas têm sido relevantes no avanço da compreensão da influência que agentes de IA têm no processo de aprendizagem em ambientes educacionais.

O estudo [22] se aprofundou no conceito de agentes pedagógicos (AP): agentes utilizados em ambientes de aprendizado online com fins educacionais. Esses agentes desempenham diversas funções educacionais, sendo capazes de realizar simulações realistas e focadas nas necessidades socioculturais dos alunos. O trabalho fez uma revisão da literatura sobre APs para identificar, analisar e avaliar o impacto da implementação desses sistemas no campo da pedagogia, de forma à considerar os diversos elementos incorporados, como texto, voz, imagens e figuras humanas.

De forma semelhante, o estudo [23] investigou a influência dos estímulos motivacionais no sucesso do aprendizado e na motivação em um ambiente de aprendizado digital. Para isso foi implementado um AP em um ambiente de aprendizado online para apoiar a motivação dos aprendizes. Os resultados do estudo, realizado com 60 estudantes do ensino fundamental, sugerem que os aprendizes com um AP alcançam um nível mais elevado de conhecimento do que aqueles sem um AP. O estudo concluiu também que a presença do AP não apenas desencadeou processos motivacionais, mas também apoiou os alunos na reflexão e elaboração dos conteúdos, resultando em uma menor carga cognitiva percebida.

O estudo [24], de 2023, faz uma análise do uso de *chatbots* impulsionados por IA para aprimorar experiências de aprendizado e de engajamento no estudo de ciência da computação. Para isso os pesquisadores desenvolveram um ambiente de aprendizado com a utilização de quatro papéis distintos de *chatbot*: Instrutor, Companheiro, Orientador de Carreira e Apoiador Emocional. Por meio de sua abordagem inovadora, o estudo fornece percepções significativas sobre o potencial dos *chatbots* com múltiplos papéis, impulsionados por IA, em remodelar o cenário da educação em ciência da computação e promover um ambiente de aprendizado envolvente.

O trabalho [25] apresentou o primeiro modelo de dinâmica de afetos utilizando dados de alunos brasileiros aprendendo com um Sistema Tutor Inteligente (STI) de matemática baseado em passos. Este modelo, denominado "pat2math"possibilitou a descoberta de transições significativas entre as emoções acadêmicas confusão, frustração, tédio e engajamento. Além disso, o trabalho apresentou uma análise da dinâmica de afetos considerando o sexo dos alunos. Por meio desta análise, o trabalho concluiu que existem transições significativas dependentes do sexo do aluno, como no caso da transição de tédio para engajamento, que só é significativa para alunos do sexo feminino.

[1]

[1]Endereço online do pat2math (2023): http://www.projeto.unisinos.br/pat2math/

Com a mesma premissa dos já citados, diversos outros trabalhos se dedicaram a analisar e coletar informações à respeito do uso de APs. O artigo [26] explora a história dos APs e resume o estado atual do conhecimento. Uma vez estabelecido o estado atual dos APs, são discutidas as barreiras e oportunidades para sugerir como os pesquisadores podem melhorar da melhor forma os agentes pedagógicos e sua implementação.

O trabalho [27] analisou a literatura produzida entre anos de 2007 a 2017 e concluiu que existem diversos tipos de design de APs e que sua utilização têm demonstrado mudanças positivas no comportamento de aprendizes e tutores. O artigo [28] analisou o funcionamento de um AP simples, desenvolvido para treinar trabalhadores em fábricas de calçados [29] e buscou aprimorar o comportamento dos APs de forma a torná-los mais comunicativos e expressivos, aumentando sua eficácia.

2.3 SMAs educacionais

O trabalho [30], aqui em destaque, propõe um modelo interativo de ensino baseado em SMA, onde foram desenvolvidos modelos de agentes, arquitetura de agentes, linguagem e protocolos de interação.

No estudo, o autor entende que a qualidade de um sistema composto por uma máquina tutora e um aprendiz humano está relacionada com a capacidade que a máquina tem de atender as necessidades específicas de cada aluno. A partir deste pressuposto, o estudo envolveu a integração das áreas de STI e SMA para a criação de um ambiente multidimensional de conhecimento. Como resultado do estudo, surgiu a definição de um sistema tutor multi-agentes e a elaboração de um mecanismo para a construção do modelo do aprendiz, por meio de uma modelagem distribuída.

Vindo do mesmo autor e com uma temática muito parecida, o trabalho [31] apresenta um ambiente computacional de aprendizado denominado Mathema. A ideia do ambiente é integrar alunos em uma sociedade de agentes tutores artificiais, visando envolvê-los em uma situação de aprendizado por meio de um processo cooperativo e interativo. Cada agente tutor oferece suporte, entre outros elementos, a características de três módulos de conhecimento fundamentais em um STI clássico, projetado para saber o que é ensinado (módulo de domínio), para quem é ensinado (modelo do aluno) e como é ensinado (módulo pedagógico). Dessa forma, o modelo busca criar um ambiente mais flexível e rico, aprimorando a adaptabilidade do sistema em relação ao aluno.

Outro estudo em destaque na mesma área, é o [32], que utilizou os conceitos de multi-agentes aplicados na computação afetiva. Segundo os autores, as emoções têm um papel relevante no aprendizado dos seres humanos, estando diretamente relacionadas as capacidades cognitivas humanas. A computação afetiva é uma área da computação que

desenvolve aplicações e produtos capazes de influenciar as emoções humanas para se atingir determinada finalidade. Neste sentido, o estudo apresenta discussões que tratam da introdução de características afetivas em projetos de SMA e STI.

O estudo traz a noção de que em um STI pode ser mais importante inferir e decidir sobre o estado emocional do aluno/usuário do que necessariamente fazer com que o sistema exprima emoções. Já em um SMA, fazer com que os agentes expressem emoções aos usuários pode ser mais relevante que inferir o estado emocional dos usuários em si.

O trabalho [33] fez um estudo sobre a prática da educação ambiental na disciplina de Matemática por meio da modelagem matemática. O trabalho apresentou e discutiu as implicações desse processo sob as perspectivas conservadora e crítica da educação ambiental e de aspectos teórico-metodológicos da modelagem matemática. A pesquisa utilizou um ambiente clássico de programação multiagente, denominado "Netlogo", que tem sido utilizado com fins educacionais. A pesquisa foi realizada com quatro turmas do 9º ano do ensino fundamental de uma escola pública de São Lourenço do Oeste, Santa Catarina e chegou a conclusão de que os estudantes apresentaram melhor compreensão da realidade ambiental e da matemática por meio da modelagem matemática.

[2]

2.4 Considerações finais do capítulo

Dado o cenário descrito e as análises feitas pelos trabalhos mencionados torna-se evidente a importância que assistentes virtuais podem desempenhar durante o aprendizado em ambientes educacionais, trazendo benefícios aos aprendizes e tutores. Diante disso este trabalho se propõe a realizar uma análise, do ponte de vista ético, dos possíveis comportamentos dos agentes de um modelo de SMA do SmartUnB.ECOS, com o intuito de que se haja uma compreensão prévia dos possíveis impactos que esse sistema pode ter nos usuários antes que ele seja aplicado na prática.

[2]Endereço online do Netlogo (2023): https://ccl.northwestern.edu/netlogo/

Capítulo 3

Fundamentação para a proposta

Este capítulo será dividido em 4 seções. A primeira irá apresentar trabalhos relacionados às discussões éticas no desenvolvimento de IA's. A segunda irá abordar o conceito de sociedades multi-agentes e um meta-modelo útil para desenvolvê-las. Na terceira serão apresentados estudos realizados por alunos da UnB, com foco no desenvolvimento do projeto SmartUnB.ECOS, analisando o atual cenário em que se encontra o projeto. Ao final do capítulo pretende-se entender como este estudo pode ser útil às questões éticas que envolvem o ecossistema SmartUnB.ECOS.

3.1 Arcabouço adotado para análise da ética computacional

Com a popularização das tecnologias de IA, algumas questões polêmicas têm sido levantadas, como por exemplo a situação que ocorreu durante o recrutamento de recursos humanos na empresa de tecnologia Amazon, onde a ferramenta de IA usada tornou-se preconceituosa em relação à contratação de mulheres [34], ou também o caso em que um carro autônomo da empresa Tesla bateu em uma van enquanto estava no modo de piloto automático [35].

Para lidar com essas questões, alguns trabalhos na área da computação têm sido realizados com o objetivo de analisar quais os principais princípios éticos que devem ser considerados durante a fase de desenvolvimento de sistemas que utilizam IA.

3.1.1 Princípios éticos de [1]

O artigo "Artificial Intelligence: the global landscape of ethics guidelines"[1], descreve o trabalho realizado por pesquisadores que coletaram e analisaram dados a respeito das inúmeras diretrizes existentes no campo da ética em IA, e com base na frequência com a

qual as diretrizes aparecem, os autores apontaram onze princípios considerados como os mais relevantes a serem observados durante o desenvolvimento de um sistema IA.

O artigo "The Current State of Industrial Practice in Artificial Intelligence Ethics", publicado pela IEEE em 2020 [7], analisou dados de 211 empresas desenvolvedoras de softwares, também com o objetivo de descobrir quais princípios éticos têm sido seguidos com mais frequência por elas e identificou os mesmos princípios elencados em [1]. Os princípios identificados são:

- Transparência.
- Justiça e equidade.
- Não-maleficência.
- Responsabilidade.
- Privacidade.
- Beneficência.
- Liberdade e autonomia.
- Confiança.
- Sustentabilidade.
- Dignidade.
- Solidariedade.

O estudo de José Antônio Siqueira explorou o significado de cada um dos princípios no contexto da computação [36]. Abaixo estão elencados e contextualizados os princípios tendo como base os três estudos considerados [1, 7, 36].

Transparência

Um dos princípios mais debatidos e relevantes por ser aquele que permite que os demais tenham visibilidade. É essencial que exista transparência nos projetos de desenvolvedores de IA e que eles estejam alinhados à legislação atual, que no caso do Brasil inclui: a Lei Geral de Proteção de Dados (LGPD) [1] e o Regulamento Geral de Proteção de Dados (GDPR) [2].

[1] Lei no. 13.709 de 14 de agosto de 2018 - Lei Geral de Proteção de Dados Pessoais (LGPD): Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural [37].

[2] O Regulamento Geral de Proteção de Dados é, um regulamento de dados da União Europeia (UE), que tem efeitos sobre o Brasil, devido ao fluxo de dados, principalmente aquele relacionado ao comércio online, que força os comerciantes nacionais a se adaptarem à legislação em vigor na UE [38].

Justiça e equidade

Esses princípios têm sido debatido na mídia e na academia por tratarem-se de questões de discriminação e injustiça, principalmente em relação a grupos minoritários e de pouca representatividade social. Os desenvolvedores devem se preocupar em criar sistemas plurais, e que garantam a inclusão desses grupos, para evitar promover algum tipo de exclusão social.

Não-maleficência

Os impactos negativos dos sistemas de IA aos seus usuários, devem ser antecipados e evitados pelos desenvolvedores, para tal se faz necessário a testagem prévia dos sistemas, antes de sua disponibilização, e que as organizações responsáveis pelos sistemas de IA, desenvolvam sistemas de segurança contra possíveis ataques.

Responsabilidade

Em virtude da autonomia dos sistemas de IA, é importante que eles possuam mecanismos que permitam identificar os responsáveis pelos casos de mal funcionamento, vazamento de dados ou de violações de segurança.

Privacidade

Este é o princípio mais prevalente após o surgimento das leis LGPD e GDPR. A preservação da privacidade dos usuários é fundamental, em especial, durante a manipulação das quantidades massivas de dados que os sistemas IA utilizam para funcionar.

Beneficência

O desenvolvimento de sistemas IA devem ter como intento beneficiar a sociedade como um todo, levando em conta não só o bem estar dos indivíduos, como também a preservação do meio ambiente.

Liberdade e autonomia

Os sistemas baseados em IA não devem comprometer a liberdade e a autonomia dos indivíduos. Os usuários não podem ser manipulados por esses sistemas e as organizações devem preservar a liberdade de expressão dos indivíduos. Toda e qualquer utilização de dados dos usuários deve ser por eles consentida.

Confiança

Organizações desenvolvedoras de IA devem buscar a confiança de seus usuários, agindo com transparência [3.1.1] e responsabilidade [3.1.1], e demonstrando que seus sistemas são inofensivos e benevolentes.

Sustentabilidade

Organizações de sistemas IA devem buscar ser ecologicamente sustentáveis, prezando a eficiência energética e o baixo consumo de recursos naturais, especialmente nestes tempos em que as mudanças climáticas tem sido amplamente discutidas e ações para minimizá-las tem sido elaboradas.

Dignidade

Sistemas baseados em IA devem procurar entender e respeitar os valores humanos, suas culturas e seus direitos, em busca de não violá-los. Deve ficar sempre claro para o usuário, que sua interação, é com uma máquina, e não com um outro ser humano.

Solidariedade

Este princípio se relaciona ao desenvolvimento humano e as relações sociais como um todo. As organizações de sistemas IA devem promover a coesão social, apoiando os sistemas democráticos e o desenvolvimento humano. Além disso, os desenvolvedores devem trabalhar no sentido de evitar a criação de sistemas que facilitem a propagação de fake news ou que permitam a violação da privacidade dos usuários.

3.1.2 Matriz de riscos e benefícios de [2]

Rangel em seu estudo de 2021, analisou a ética de alguns aspectos da rede CICFriend [2]. Ao longo de seu trabalho ela desenvolveu, entre outras coisas, uma matriz de riscos e benefícios referentes às operações de tratamento de dados dos usuários da rede. A matriz é feita de linhas contendo o tipo de tratamento a ser analisado e colunas de riscos e benefícios para o tratamento em questão.

Os tipos de tratamentos escolhidos pela autora correspondem às fases do ciclo de vida do tratamento de dados pessoais, evidenciados pelo próprio guia de boas práticas da LGPD [39]: coleta, armazenamento, processamento, apresentação e compartilhamento. A matriz de Rangel em detalhes encontra-se ilustrada no Apêndice A.

3.2 Meta-modelo para modelagem da SMA

Os algoritmos de inteligência artificial distribuída (IAD) são classificados em três categorias, cada uma delas referentes aos métodos utilizados para solucionar suas respectivas tarefas. As categorias em questão são: IA paralelas, solução distribuída de problemas (SDP) e sistemas multi-agentes (SMA) [17].

As IA paralelas funcionam, por meio do aproveitamento do paralelismo das tarefas, desenvolvendo algoritmos e arquiteturas que tornam os algoritmos de IA em padrões mais eficientes.

O método de SDP consiste em dividir uma tarefa em subtarefas, onde cada uma delas está ligada à um nó contido em um conjunto de nós cooperativos, conhecidos como entidades computacionais. Essas entidades compartilham informações entre si e também entre outras entidades, o que acaba limitando sua flexibilidade.

Os SMA, foco deste trabalho, consistem em entidades autônomas denominadas agentes. Apesar de parecidos com as entidades de SDP, os agentes apresentam maior flexibilidade em função de sua capacidade de aprender e tomar decisões autônomas.

3.2.1 Método Aalaadin e meta-modelo AGR

O método Aalaadin, de Jacques Ferber e Olivier Gutknecht [3] é um meta-modelo genérico para construção de sociedades multi-agentes com base no modelo AGR. O termo AGR se refere aos três principais conceitos utilizados nesse tipo de modelo: Agentes, grupos e papéis [40]. No Aalaadin esses conceitos funcionam da seguinte forma: o agente representa qualquer entidade comunicativa ativa e pode assumir diferentes papéis em diferentes grupos. Um grupo é um conjunto de agentes agregados. Cada agente pode fazer parte de um ou mais grupos. Um grupo pode ser fundado por um agente e os agentes devem pedir permissão uns aos outros para se juntar à um grupo já existente. Por fim, um papel é uma representação abstrata da função de um agente. Um agente pode possuir diversos papéis e cada papel pertence à um respectivo grupo (verificar figura 3.1).

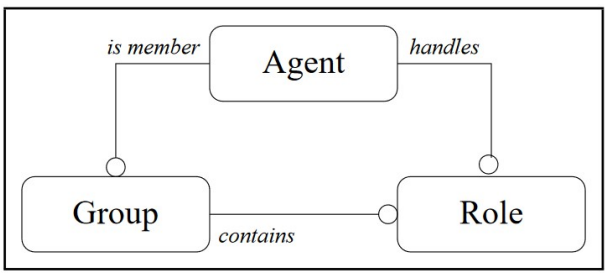

Figura 3.1: Estrutura básica AGR (retirado de [3])

Em cada grupo deve haver um agente com o papel especial de gerente do grupo. Esse papel gerencial é destinado automaticamente ao agente criador do grupo, responsável por lidar com as permissões relativas aos papéis e com as requisições de entrada e saída do grupo. Ele também pode revogar papéis e remover agentes do grupo caso seja necessário.

Caso possuam a capacidade de serem serializáveis, os agentes têm a possibilidade de migrar de um grupo para outro. Cada conjunto de grupos locais possui um grupo chamado *grupo de mobilidade* que contém todos os *itinerantes*, papel atribuído aos agentes que desejam migrar. Dessa forma um único agente com o papel de *migratório* no grupo de mobilidade dá início ao processo de migração. Durante o processo tanto o grupo de origem quanto o grupo de destino do agente itinerante devem se comunicar, certificando-se que a migração é possível. Agentes que não podem ser serializados não podem ser aceitos como itinerantes.

Além dos conceitos centrais do modelo AGR há também outros conceitos, que não são representados diretamente em organizações multiagentes, que servem apenas como ferramentas de análise e design (figura 3.2).

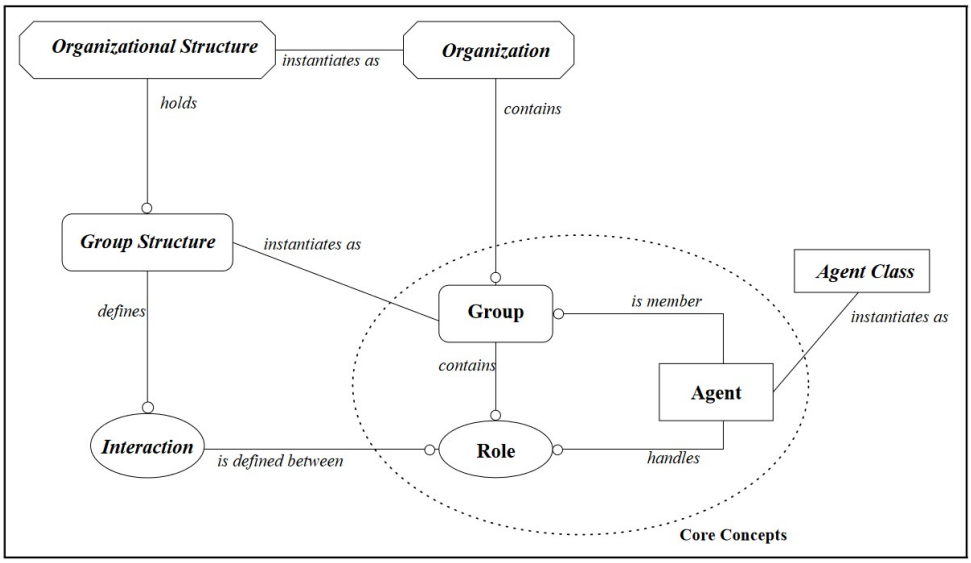

Figura 3.2: Modelo metodológico adicional (retirado de [3])

3.2.2 Biblioteca Java MaDKit de apoio à implementação

Os autores de [3] construíram a ferramenta "MadKit"[41], uma plataforma de agentes em linguagem Java, feita com o objetivo de se utilizar o meta-modelo Aalaadin na prática. A plataforma implementa na prática os conceitos de agentes, grupos e papéis, e adiciona três princípios de design: arquitetura Micro-Kernel, agentificação dos serviços e modelo de componente gráfico. A ideia da plataforma MadKit é ser usada de forma adaptativa

para diferentes situações e projetos, e pode ser utilizado para se implementar o modelo AGR proposto neste trabalho (Figura 3.3).

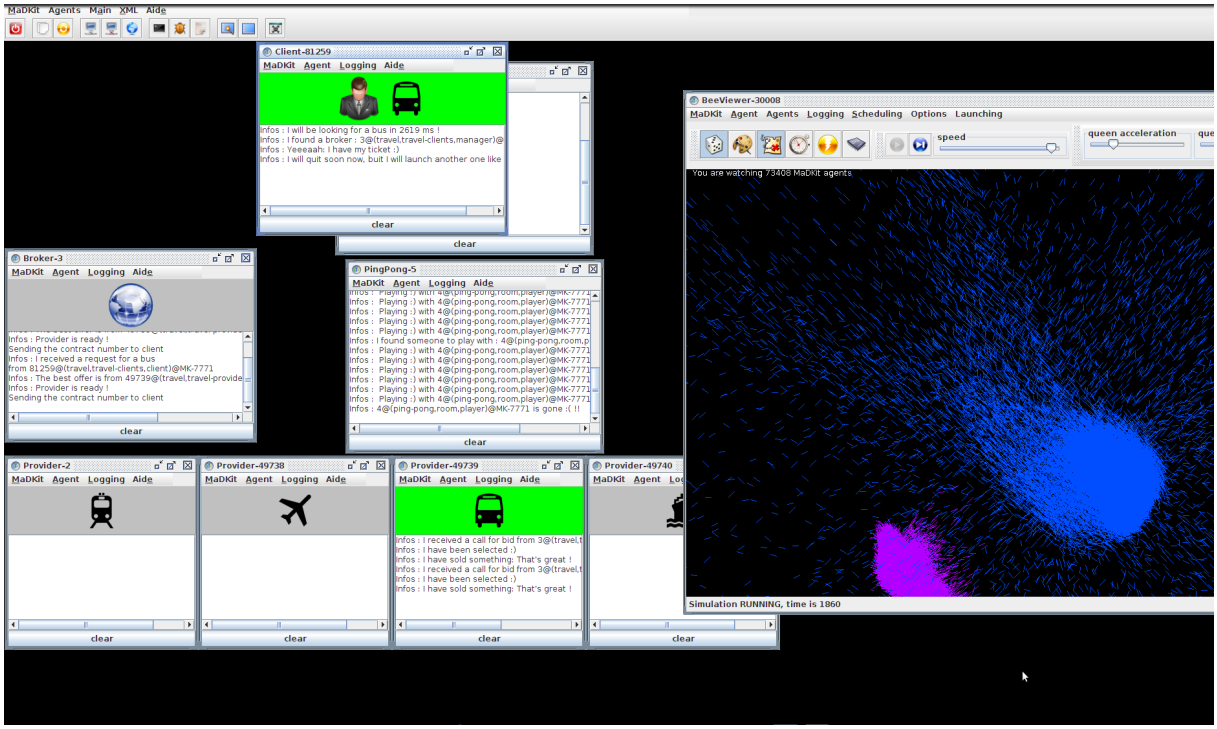

Figura 3.3: Exemplo do software Madkit em execução

[3]

3.2.3 AGR4BS: atualizações recentes do meta-modelo

Com a ascensão das tecnologias de blockchain, atrativas por permitirem registros ordenados e auditáveis de transações acessíveis à qualquer pessoa, o trabalho [42] desenvolveu um modelo organizacional de multi-agentes, chamado "AGR4BS", com o objetivo de se estudar sistemas de blockchain. O trabalho citado propõe utilizar o modelo AGR desenvolvido para identificar e representar as entidades genéricas desse tipo de sistemas.

3.3 Projeto smartUnB.ECOS

O SmartUnB.ECOS consiste em um ecossistema educacional digital, desenvolvido na UnB, que tem como objetivo atender a comunidade do campus universitário, fomentando a socialização e a aprendizagem por meio da interoperabilidade de ferramentas de aprendizagem e de comunicação. O ecossistema é proposto para servir aos docentes e discentes do Departamento de Ciência da Computação da UnB [4].

[3]Endereço online oficial do Madkit (2023): http://www.madkit.net/madkit/index.php

A ideia do SmartUnB.ECOS é que os próprios alunos possam contribuir para a construção e o aperfeiçoamento do projeto, desenvolvendo e aperfeiçoando diferentes elementos do ecossistema.

3.3.1 Elemento *Kirilândia*

O ecossistema está atualmente estruturado em nove elementos sobre os quais se busca estimular uma dinâmica entre aprendizado formal e informal. Esses elementos estão ilustrados na Figura 3.4, mas para este estudo tem-se como foco apenas um deles, a sociedade multi-agentes (elemento de número oito na figura), explicada na seção anterior, e denominada *Kirilândia*.

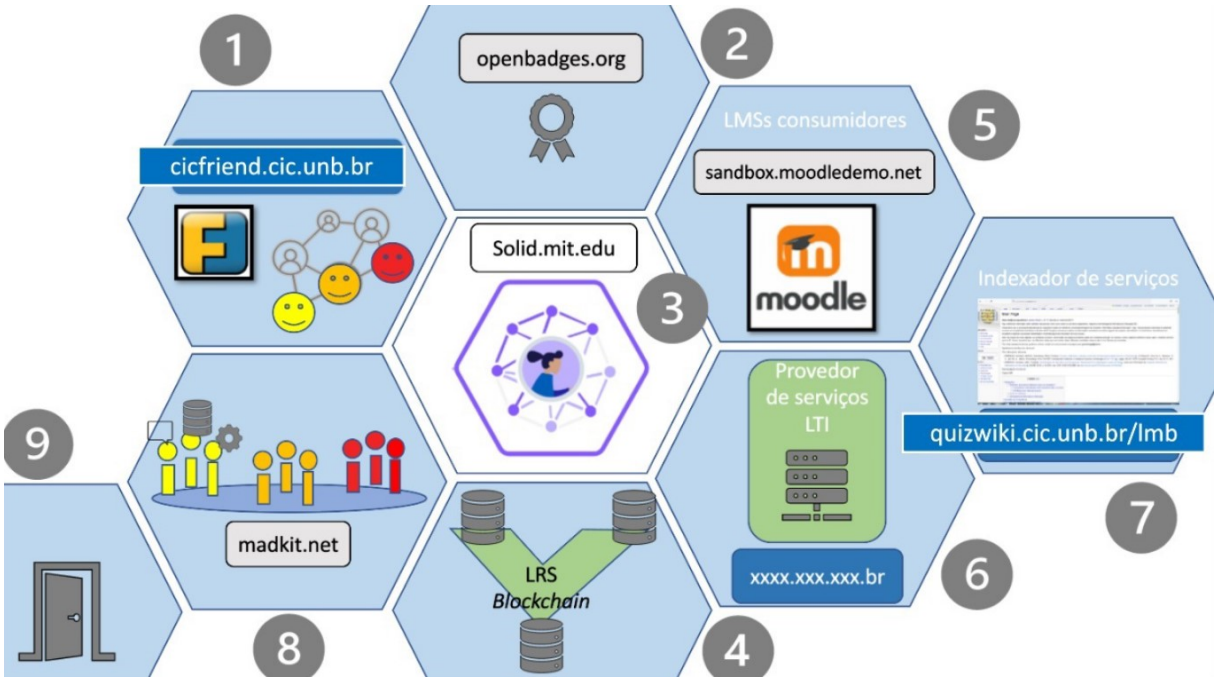

Figura 3.4: SmartUnB.ECOS [4]

3.3.2 Acoplamento com outros trabalhos na equipe

Como mencionado em 1.2 este trabalho é parte integrante do projeto SmartUnB.ECOS e se encontra diante de vários outros trabalhos que buscaram, assim como este, desenvolver a *Kirilândia* e o projeto SmartUnB.ECOS como um todo. Alguns destes trabalhos dialogam diretamente com o tema aqui abordado.

No trabalho de TCC [43], Claiton Custódio se aprofundou na modelagem de conhecimento dos agentes da *Kirilândia*, propondo um modelo de estudante holístico, no formato de uma ontologia, voltado para o estudante de graduação do Departamento de Ciência

da Computação da UnB. A utilização de um modelo comum, que seja utilizado pelas diversas aplicações e serviços do ecossistema educacional do SmartUnB.ECOS, possibilita o compartilhamento de dados e a integração das aplicações.

Outro trabalho que possui conexão direta com este é o TCC [44], em que Thiaggo Ferreira e Oscar Etcheaverry analisaram a implementação de um POD server no ecossistema do SmartUnB.ECOS e a introdução dos alunos ao uso dos Solid PODs, com foco na descentralização da web. O trabalho chegou à conclusão de que a utilização do POD server pode trazer benefícios para o aprendizado dos alunos, dando-lhes maior controle sobre seus dados e permitindo a criação de um ambiente mais colaborativo e personalizado.

Ainda no contexto de armazenamento e controle de acesso aos dados, o TCC do João Duda [45] também analisou a implementação da tecnologia Solid no ecossistema do SmartUnB.ECOS. Para isso os autores do trabalho mapearam algumas aplicações Solid-compatíveis e desenvolveram duas aplicações próprias com o objetivo de avaliar como seriam suas implementações no SmartUnB.ECOS. O trabalho concluiu que, apesar de ainda carecer de divulgação para que se torne mais popular, atualmente já é possível desenvolver projetos consistentes com Solid.

3.4 Considerações

Cabe destaque, que com o desenvolvimento da *Kirilândia* e do SmartUnb, surgem preocupações relacionadas com a segurança dos usuários e com as questões éticas do ecossistema, tornando-se necessário uma análise detalhada que considere as possíveis falhas e as questões sensíveis que possam inibir os usuários ou até mesmo comprometer sua experiência de uso e suas informações pessoais.

Para a realização da análise, o presente trabalho apresentará uma proposta de modelo AGR para a *Kirilândia*, que permita analisar as suas diversas etapas de funcionamento e as possíveis interações entre seus agentes autônomos. Serão usadas como base as fases do tratamento de dados pessoais da matriz de riscos e benefícios de Rangel [2] e os onze princípios éticos considerados do estudo [1].

Capítulo 4

Proposta: contribuições para um projeto SMA eticamente consciente

Este capítulo se divide em duas seções, na primeira seção buscou-se apresentar uma possível descrição de sociedade autônoma para a Kirilândia segundo o modelo AGR. Na segunda seção foi elaborada uma análise ética do que foi descrito usando-se como base a matriz de riscos e benefícios do estudo de Rangel [2] seguida da análise da mesma descrição sob o ponto de vista dos onze princípios éticos do estudo [1], buscando viabilizar a análise ética de cada uma das possíveis interações entre agentes e usuários, identificando os principais riscos dessas interações e quais os princípios éticos mais relevantes para a construção da CICFriend.

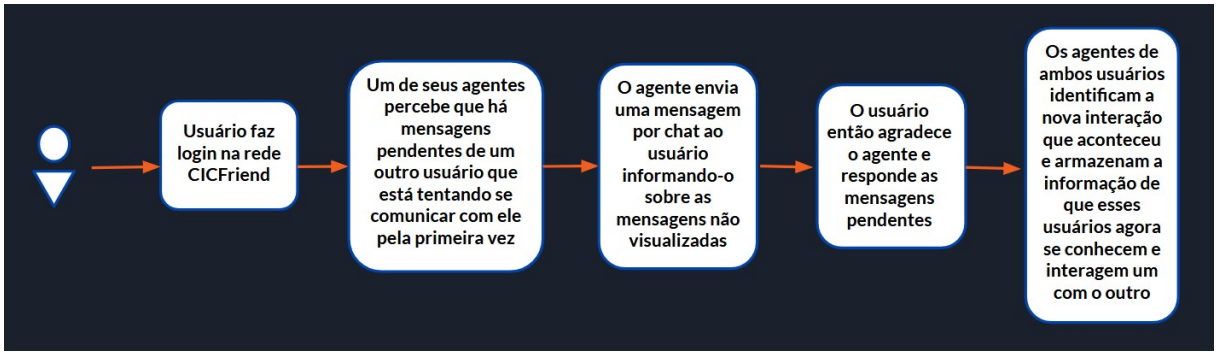

Figura 4.1: Exemplo de atuação dos agentes autônomos na rede CICFriend

4.1 Projeto da sociedade segundo modelo AGR

Para ilustrar uma das possíveis atuações dos agentes autônomos a figura 4.1 apresenta uma interação hipotética que poderia acontecer entre dois usuários na rede CICFriend. Tendo por base o método Aalaadin [3], pode-se dar início ao protótipo do modelo de

agentes da *Kirilândia*. Para tal inicia-se definindo a estrutura de grupos, que consiste na criação dos grupos necessários, com seus respectivos papéis. Em seguida é apresentada a estrutura organizacional, definindo-se a forma como os grupos se relacionam entre si.

4.1.1 Estrutura de grupos

Para este projeto pode-se ter 3 grupos principais: Interface, conhecimento e raciocínio. A estrutura de um grupo é definida como sendo uma tupla $S = \langle R, G, L \rangle$, onde:

- R é um conjunto finito de possíveis papéis que um agente pode interpretar dentro do grupo. $R = \{r_1, r_2, r_3, ...\}$.
- G é um grafo orientado contendo as possíveis interações entre os diferentes papéis. $G : R \times R \rightarrow l$. A orientação de cada aresta indica qual papel está iniciando a interação. Cada aresta representa uma interação com seu respectivo rótulo l.
- L é a linguagem de interação. Para cada interação do grafo G é escolhido um protocolo formal p de interação entre os papéis contidos em G. $\forall (r_i, r_j, l) \in G, \exists! p \in L$.

Cada grupo está descrito nas subseções abaixo.

Interface

O grupo de interface é aquele que contém os agentes cujo papel se relaciona com a comunicação entre agentes e usuários e esses com outros usuários, que podem ser tanto discentes quanto docentes. Este grupo poderá enviar informações para o grupo de conhecimento caso necessário, e para realizar sua função, esse grupo deve conter, além do papel padrão de gerente, os seguintes papéis: agentes de chat e agentes de notificações.

Os agentes de chat irão se comunicar com os usuários através de mensagens personalizadas e, se necessário e com o devido cuidado, monitorar alguns tipos de interações entre eles. Os agentes de notificações são responsáveis por informar notícias, enviar alertas e sugestões aos usuários. Tanto os agentes de chat, quanto os agentes de notificações podem se comunicar com o agente gerente para solicitar uma troca de seu papel por outro papel disponível no grupo.

Para este grupo tem-se a seguinte tupla: $S_i = \langle R_i, G_i, L_i \rangle$.

- $R_i = \{r_c, r_n, r_g\}$, sendo r_c o papel de chat, r_n o papel de notificações e r_g o papel de gerente do grupo.
- $G_i : (r_c \times r_g \leftrightarrow l_{cg}), (r_n \times r_g \leftrightarrow l_{ng})$.
- $\forall (e \in E(G_i)) \exists! p \in L_i$.

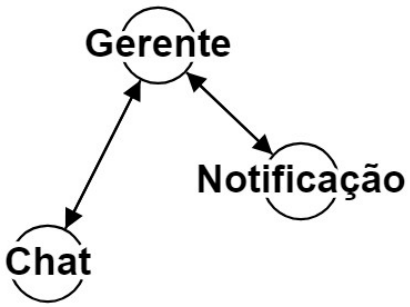

Figura 4.2: **Grupo de Interface**

Conhecimento

O grupo de conhecimento cuida da coleta e produção de todos os dados dos usuários que aceitarem fornecer seus dados à CICFriend. Esses dados podem incluir informações pessoais como:

1. Informações de Perfil:
    - Nome
    - Idade
    - Fotos
    - Gênero
    - Período que está cursando
    - Hobbies e áreas de interesse
2. Comportamentos de navegação:
    - Conteúdos visualizados na plataforma
    - Tempo gasto em diferentes sessões da rede social
    - Conteúdo interagido
3. Conexões com outros usuários:
    - Usuários com quem interage
    - Históricos de mensagens privadas
4. Informações de dispositivo:
    - Tipo de dispositivo utilizado (computador, smartphone, tablet)
    - Sistema operacional

- Endereço IP

5. Informações de registro e atividades
    - Horários de login e logout
    - Alterações no perfil
    - Atividades recentes na plataforma
6. Configurações de privacidade
    - Preferências de privacidade e segurança
    - Configurações de visibilidade do perfil
    - Histórico de ajustes de privacidade
7. Entre outros

Para realizar sua função, este grupo deve conter, além do papel padrão de gerente, os seguintes papéis: agentes de coleta de dados e agentes de eliminação de dados.

Os agentes de coleta de dados são responsáveis por coletar e armazenar os dados dos usuários, sejam eles dados informados pelos próprios usuários, ou dados internos do sistema da UnB aos quais a SMA tenha acesso. Os agentes de eliminação de dados são encarregados de se desfazer dos dados que não tenham mais serventia para o sistema, ou cuja remoção tenha sido solicitada pelos usuários. Tanto os agentes de coleta, quanto os agentes de eliminação podem se comunicar com o agente gerente para solicitar uma troca de seu papel por outro papel disponível no grupo.

Para este grupo tem-se a seguinte tupla: $S_c = \langle R_c, G_c, L_c \rangle$.

- $R_c = \{r_c, r_e, r_g\}$, sendo r_c o papel de coleta de dados, r_e o papel de eliminação de dados e r_g o papel de gerente do grupo.
- $G_c : (r_c \times r_g \leftrightarrow l_{cg}), (r_e \times r_g \leftrightarrow l_{eg})$.
- $\forall(e \in E(G_c))\exists!p \in L_c$.

Raciocínio

O grupo de raciocínio manipula os dados coletados e armazenados pelo grupo de conhecimento. Este grupo gerencia todas as possíveis conexões entre usuários, assim como as relações entre projetos e atividades similares. Este grupo deve conter, além do papel padrão de gerente, os seguintes papéis: agentes de relacionamentos e agentes de projetos.

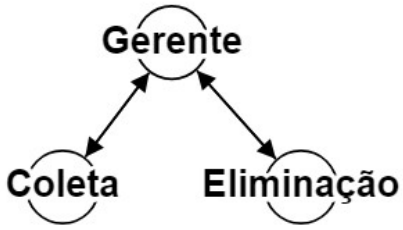

Figura 4.3: **Grupo de Conhecimento**

Os agentes de relacionamentos analisam os dados individuais de cada usuário, coletados pelo grupo de conhecimento, e realizam comparações entre eles com base em critérios pré-determinados, como perfil pessoal por exemplo, de forma a criar novas conexões entre os usuários da rede e gerenciar conexões já existentes. Da mesma forma, os agentes de projetos podem analisar características específicas de projetos e atividades dos usuários como: objetivo do estudo, alunos e professores participantes, área de atuação, entre outros, e catalogá-los de diferentes formas.

Os agentes de relacionamentos e os agentes de projetos podem enviar informações para os agentes do grupo de interface, para que estes por sua vez, comuniquem certas decisões tomadas no grupo de raciocínio, por meio de sugestões e notificações, aos usuários da rede.

Para este grupo tem-se a seguinte tupla: $S_r = \langle R_r, G_r, L_r \rangle$.

- $R_r = \{r_r, r_p, r_g\}$, sendo r_r o papel de relacionamentos, r_p o papel de projetos e r_g o papel de gerente do grupo.
- $G_r : (r_r \times r_g \leftrightarrow l_{rg}), (r_p \times r_g \leftrightarrow l_{pg})$.
- $\forall (e \in E(G_r)) \exists ! p \in L_r$.

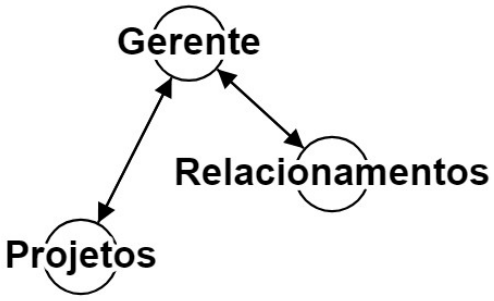

Figura 4.4: **Grupo de Raciocínio**

4.1.2 Estrutura organizacional

A estrutura organizacional é definida como sendo uma dupla $O = \langle S, Rep \rangle$, em que:

- S é um conjunto de estruturas de grupos válidas.
- Rep é um grafo representativo, em que cada aresta entre dois grupos $S_a S_b$ representa uma interação entre dois papéis $r_1 r_2$, sendo $r_1 \in R_1$ e $r_b \in R_2$, com $S_a \in S$, $S_b \in S$, $R_1 \in S_a$, $R_2 \in S_b$.

Para simplificar o modelo e facilitar a sua visualização, os três grupos podem compartilhar o mesmo agente gerente. Dessa forma:

- $S = \{S_i, S_c, S_r\}$.
- $Rep : (r_{projeto} \times r_{notificação} \rightarrow l_{pn}), (r_{projeto} \times r_{chat} \rightarrow l_{pc}),$
  $(r_{projeto} \times r_{coleta} \leftrightarrow l_{pc}), (r_{projeto} \times r_{eliminação} \leftrightarrow l_{pe}),$
  $(r_{relacionamento} \times r_{notificação} \rightarrow l_{rn}), (r_{relacionamento} \times r_{chat} \rightarrow l_{rc}),$
  $(r_{relacionamento} \times r_{coleta} \leftrightarrow l_{rc}), (r_{relacionamento} \times r_{eliminação} \leftrightarrow l_{re}),$
  $(r_{coleta} \times r_{notificação} \rightarrow l_{cn}), (r_{coleta} \times r_{chat} \leftrightarrow l_{cc}).$

A estrutura organizacional do modelo definido está representada na Figura 4.4.

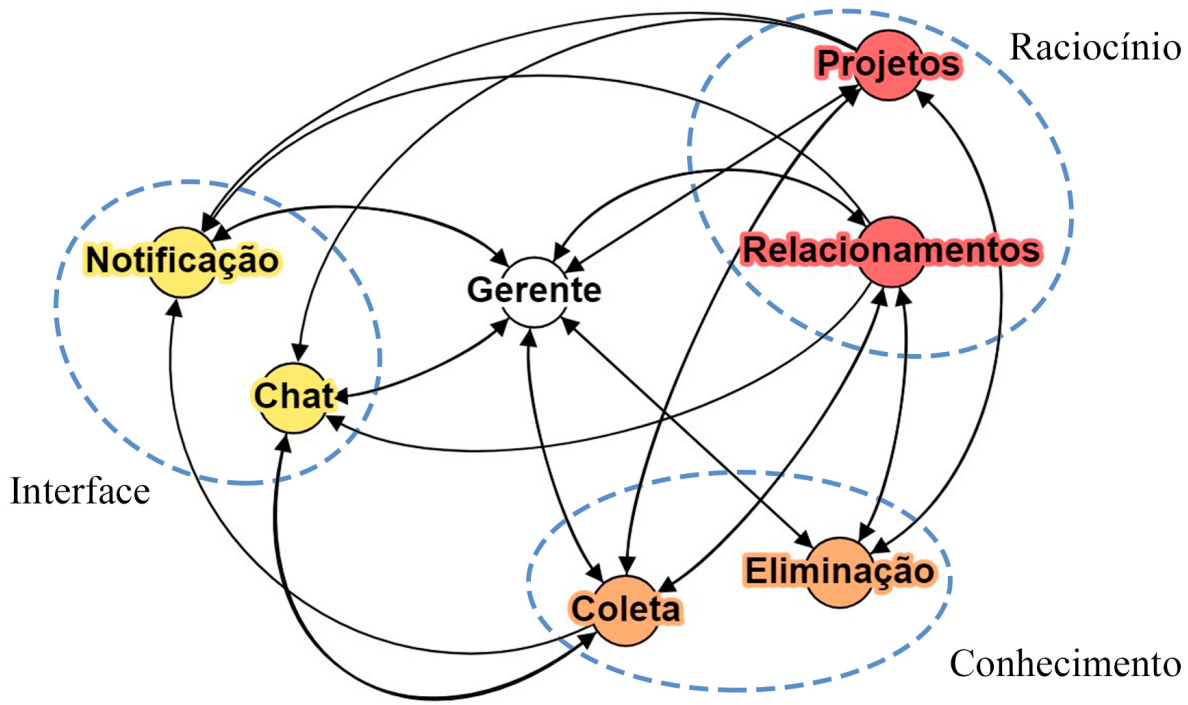

Figura 4.5: **Estrutura organizacional**

4.2 Análise pela matriz de riscos e benefícios

A partir do que foi descrito é possível traçar um paralelo entre as funcionalidades e as etapas de tratamento dos dados abordadas pela matriz de Rangel [2]. No caso do modelo aqui proposto, os dados em questão são as informações, pessoais e acadêmicas, dos usuários da rede CICFriend.

É possível perceber que todas as etapas em questão encontram-se espalhadas ao longo do modelo AGR proposto. As etapas de coleta, retenção e eliminação de dados, apontadas no estudo de Rangel, estão inclusas nas funcionalidades relativas aos papéis dos agentes do grupo de conhecimento do modelo aqui posto, as etapas de processamento de dados estão presentes no grupo de raciocínio, e as etapas de compartilhamento de dados estão presentes, tanto no grupo de raciocínio, quanto no grupo de interface.

Essas cinco etapas de manipulação de dados possuem riscos que precisam ser levados em consideração durante o desenvolvimento da rede. Tendo isso em mente e usando como base as informações da matriz de Rangel, é possível tecer algumas considerações.

4.2.1 Grupo de interface

Por receberem informações individuais dos usuários e encaminhá-las ao grupo de conhecimento e à outros usuários, os agentes do grupo de interface estão propensos a apresentar os seguintes riscos:

- Durante a elaboração de novos alertas e sugestões, feitos pelos agentes notificadores, há o risco de que sejam realizadas entregas desnecessárias ou incorretas de notificações aos usuários da rede, resultando em confusões e desentendimentos por parte dos mesmos.
- Durante o monitoramento de conversas entre usuários, relativo ao papel dos agentes de chat, há o risco de que conversas e interações sejam monitoradas sem o devido consentimento dos usuários, o que resultaria em violações de sigilo e privacidade.
- Durante interações entre agentes e usuários, relativas ao papel dos agentes de chat, existe o risco de ocorrer interações em que os usuários não tenham ciência de que estão se comunicando com uma máquina e acreditem estar interagindo com outro ser humano.

4.2.2 Grupo de conhecimento

Dado que os agentes do grupo de conhecimento do modelo em estudo são responsáveis por coletar, armazenar e remover informações particulares dos usuários, esse grupo pode apresentar os seguintes riscos:

- Durante a coleta e retenção de dados, referente ao papel de coletor, há o risco de, devido à falta de informações ou engajamento dos usuários, possa ocorrer o não consentimento ou um consentimento inválido por parte dos mesmos;
- Durante a eliminação de dados, referente ao papel de eliminador, pode haver o risco da remoção indevida de informações, ocasionando perdas que prejudiquem as interações entre os usuários da rede ou o funcionamento correto da mesma.

4.2.3 Grupo de raciocínio

Por processarem as informações individuais dos usuários e encaminhá-las ao grupo de interface, os agentes do grupo de raciocínio estão propensos a apresentar os seguintes riscos:

- Durante o processamento das informações, referente aos papéis de relacionamentos e projetos, há o risco de que um possível enviesamento das análises possa levar à falta de bonificação ou à bonificação indevida dos usuários, além de possíveis manipulações das informações sem o devido consentimento;
- Durante o compartilhamento de informações individuais, também referente aos papéis de relacionamentos e projetos, há a possibilidade de ocorrer exposições desnecessárias e inadequadas, que resultariam em violações de sigilo e privacidade.

4.3 Análise pelos onze princípios éticos

Através da apresentação do modelo proposto, torna-se possível estabelecer uma correlação entre as funcionalidades de cada grupo do modelo com os princípios elencados: transparência; justiça e equidade; não-maleficência; responsabilidade; privacidade; beneficência; liberdade e autonomia; confiança; sustentabilidade; dignidade e solidariedade. Nesse contexto, é viável identificar quais desses princípios se destacam de forma mais proeminente em relação a cada grupo.

4.3.1 Grupo de interface

O grupo de interface contem os papéis de agentes de chat e de agentes de notificações. Como esse grupo é responsável por realizar interações diretamente com os usuários da rede, há três princípios que imediatamente se destacam:

- **Dignidade**: princípio essencial para que o usuário saiba que está interagindo com uma máquina e não com outro usuário, além de garantir que os agentes respeitem os valores humanos, sem ofender ou menosprezar seus direitos e culturas.
- **Solidariedade**: diretamente relacionado ao princípio da dignidade, se faz necessário para que as relações sociais e o desenvolvimento humano não sejam prejudicados.
- **Beneficência**: relacionado ao bem-estar dos usuários da rede. Necessário para que a experiência do usuário com o sistema e com outros usuários seja agradável para o mesmo.

4.3.2 Grupo de conhecimento

No grupo de conhecimento tem-se os papéis referentes a coleta e eliminação de informações pessoais dos usuários. Nesse contexto há três princípios que vem à tona:

- **Transparência**: essencial para garantir que toda e qualquer coleta e manipulação das informações dos usuários seja feita com o consentimento dos mesmos.
- **Privacidade**: relacionado com a capacidade do sistema de proteger, e se necessário encriptar, as informações pessoais dos usuários, de forma a evitar que hajam vazamentos ou compartilhamentos indevidos delas.
- **Confiança**: princípio que se faz necessário para que a rede transmita confiança aos usuários, demonstrando a eles, por meio da transparência, que suas informações pessoais estarão armazenadas de forma segura e privativa.

4.3.3 Grupo de raciocínio

O grupo de raciocínio contém os papéis de agentes de relacionamentos e agentes de projetos. Como esse grupo está responsável pela manipulação das informações dos usuários, vindas de outros grupos e de praticamente todas as lógicas internas que estão por trás dos relacionamentos entre eles, os princípios que se destacam neste contexto são:

- **Responsabilidade**: princípio que traz a preocupação em garantir que a rede tenha os mecanismos necessários para lidar com possíveis problemas de mal funcionamento, de forma a permitir identificar a sua origem e o responsável, caso exista.
- **Não-maleficência**: essencial para que as informações pessoais dos usuários sejam manipuladas de forma que não haja, para eles, nenhum tipo de exclusão ou prejuízo.
- **Justiça e equidade**: semelhante ao princípio de não-maleficência, busca garantir que as informações relativas aos usuários sejam processadas de forma equivalente e justa, sem que hajam bonificações ou prejuízos indevidos e desproporcionais.
- **Transparência**: todo o processo de manipulação de informações dos usuários deve ser transparente, de forma que eles possam estar cientes das razões pelas quais seus dados pessoais estão sendo utilizados.

Capítulo 5

Considerações Finais

Estudar a relação entre as questões éticas e o desenvolvimento das IA é importante para aperfeiçoar suas tecnologias, garantindo mais segurança na manipulação dos dados dos usuários. Se tais ferramentas forem desenvolvidas com atenção e responsabilidade, ambientes educacionais podem se beneficiar da utilização das mesmas, automatizando procedimentos e promovendo contatos entre dicentes e docentes.

5.1 Objetivos alcançados

No presente estudo foi possível contextualizar o problema atual relacionado a ética em IA, entender em que ponto, até a presente data, encontra-se sua discussão no meio acadêmico e analisar algumas das características da *Kirilândia*. Espera-se que as análises feitas aqui possam servir como um dos guias éticos para o desenvolvimento da rede em questão e para os próximos trabalhos que busquem se aprofundar em questões relacionadas a essa. Da mesma forma, espera-se que o que foi abordado aqui também sirva para contribuir com o entendimento das relações entre ética e tecnologia no contexto de ecossistemas educacionais como um todo.

5.1.1 Design organizacional de um SMA

Foi possível fazer uma descrição do modelo AGR proposto, assim como sua representação visual, que apesar de não pretender ser o modelo definitivo, talvez possa funcionar como ideia, de protótipo, de modelo para a *Kirilândia*. Espera-se também que o modelo definitivo a ser adotado, seja baseado no método Aalaadin, ou em métodos semelhantes, e que contenha características similares às aqui descritas.

5.1.2 Análise pela matriz de riscos e benefícios

A comparação do modelo AGR proposto com a matriz de riscos e benefícios do estudo de Rangel [2], possibilitou perceber que a matriz contempla várias das funcionalidades contidas na descrição do modelo, já que as etapas de tratamento de dados da matriz condizem com alguns dos papéis exercidos pelos agentes do modelo.

5.1.3 Análise pelos onze princípios éticos

Además a análise do modelo AGR proposto, com os onze princípios éticos, possibilitou perceber que alguns dos princípios se repetem com mais frequência que outros, como é o caso dos princípios de transparência e confiança por exemplo, e que praticamente todos eles se relacionam com as funcionalidades dos agentes do modelo, além de ter sido possível perceber entre os princípios elencados, quais são os que necessitam de maior atenção em cada parte do modelo.

5.2 Trabalhos futuros

O presente estudo está entre os vários trabalhos pertencentes ao ecossistema do SmartUnB. Assim sendo, existem ainda outras possibilidades de contribuição para o desenvolvimento do ecossistema. Fica como sugestão, o aperfeiçoamento e o desenvolvimento da parte prática da *Kirilândia*, a partir do esboço de modelo AGR aqui proposto, através da utilização de ferramentas como o MadKit [41], criada pelos mesmos criadores do método Aalaadin, que permite a implementação de modelos SMA na prática.

Fica também a sugestão de se definir protocolos de mitigação, para lidar com as possíveis violações dos termos de privacidade dos usuários, de forma a permitir que a *Kirilândia* possua ferramentas eficazes de redução de danos e prejuízos aos usuários.

Outra sugestão é definir as formas de interação entre os agentes da *Kirilândia*. Cada interação entre os agentes necessita de um protocolo único que precisa ser definido durante a implementação da SMA.

Cabe apontar também na direção da necessidade de se desenvolver ferramentas, ou utilizar ferramentas já existentes, para implementar na prática os onze princípios éticos aqui elencados, como o método ECCOLA por exemplo [9], integrando-os ao desenvolvimento da *Kirilândia* e do SmartUnB. Afinal, por mais que os desenvolvedores tenham consciência da importância da observância desses princípios, essa consciência precisa também garantir que eles sejam instituídos na prática, quando do desenvolvimento e da execução dos SMA's.

Referências

[1] Jobin, A., I. Marcello e V. Effy: *The global landscape of ai ethics guidelines.* Nature Machine Intelligence, 2019. viii, 2, 4, 10, 11, 18, 19

[2] Rangel, Barbara Varanda: *Contribuições para conduta Ética em três momentos na pesquisa em tecnologia educacional: Implantação, projeto e avaliação.* Universidade de Brasília, 2021. viii, 3, 13, 18, 19, 25, 30

[3] Ferber, Jacques e Olivier Gutknecht: *A meta-model for the analysis and design of organizations in multi-agent systems.* Université Montpellier II, 1998. x, 14, 15, 19

[4] Nóbrega, Germana M. da, Gabriel T. da Silva e Thiago VR Silva: *Um projeto estruturante para orientações de tcc em cursos de computação: Que oportunidades para ihc?* Em *Anais do XIII Workshop sobre Educação em IHC*, páginas 19–24. SBC, 2022. x, 2, 16, 17

[5] Shi, Feng, Jun Wang e Jun Shi: *Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for covid-19.* IEEE, 14, 2020. 1

[6] Srivastava, Shweta, Aditya Bisht e Neetu Narayan: *Safety and security in smart cities using artificial intelligence — a review.* IEEE, 2017. 1

[7] Vakkuri, Ville, Kai Kristian Kemell e Joni Kultanen: *The current state of industrial practice in artificial intelligence ethics.* IEEE, 37, 2020. 1, 11

[8] Kammuller, F., J. Augusto e S. Jones: *Security and privacy requirements engineering for human-centric iot systems using efriend and isabelle.* IEEE, 2017. 1

[9] Vakkuri, Ville, Kai Kristian Kemell e Pekka Abrahamsson: *Eccola - a method for implementing ethically aligned ai systems.* IEEE, 2020. 1, 30

[10] Ferguson, Rebecca: *Ethical challenges for learning analytics.* Journal of Learning Analytics, 6(3):25–30, 2019. 1

[11] Cerratto Pargman, Teresa, Cormac McGrath, Olga Viberg, Kirsty Kitto, Simon Knight e Rebecca Ferguson: *Responsible learning analytics: creating just, ethical, and caring.* Em *Companion proceedings 11th international conference on learning analytics & knowledge (LAK21)*, 2021. 1

[12] Holmes, Wayne, Kaska Porayska-Pomsta, Ken Holstein, Emma Sutherland, Toby Baker, Simon Buckingham Shum, Olga C Santos, Mercedes T Rodrigo, Mutlu Cukurova, Ig Ibert Bittencourt *et al.*: *Ethics of ai in education: Towards a*

community-wide framework. International Journal of Artificial Intelligence in Education, páginas 1–23, 2021. 1

[13] Luckin, Rosemary, Mutlu Cukurova, Carmel Kent e Benedict du Boulay: *Empowering educators to be ai-ready*. Computers and Education: Artificial Intelligence, 3:100076, 2022. 1

[14] Moor, James H.: *What is computer ethics?* Metaphilosophy, 1985. 1

[15] Smit, Koen, Martijn Zoet e John van Meerten: *A review of ai principles in practice*. Twenty-Fourth Pacific Asia Conference on Information Systems, 2020. 2

[16] Aberkane, Abdel Jaouad: *Exploring ethics in requirements engineering*. Utrecht University, 2018. 2

[17] Dorri, Ali, Salil S. Kanhere e Raja Jurdak: *Multi-agent systems: A survey*. IEEE, 2018. `https://ieeexplore.ieee.org/abstract/document/8352646`. 2, 14

[18] Oliveira, Jéssica S., Germana M. Da Nóbrega, Fernando W. Cruz e Roberta B. Oliveira: *Decentralized social network for the campus: Historical claims meet contemporary needs*. Em *2021 XVI Latin American Conference on Learning Technologies (LACLO)*, páginas 408–415. IEEE, 2021. 3

[19] Torres, Davi M., Gabriel de O. Estevam e Germana M. da Nóbrega: *Redes sociais descentralizadas na graduação em computação: Implantação, percepção discente, possibilidades*. Em *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, páginas 1344–1354. SBC, 2022. 3

[20] Wooldridge, Michael: *An Introduction to MultiAgent Systems*, volume Second edition. 2009. 3, 6

[21] Wooldridge, Michael e Nicholas Jennings: *Intelligent Agents: Theory and Practice*. 1995. 6

[22] Alfaro, Luis, Claudia Rivera, Jorge Luna-Urquizo, Elisa Castañeda, Jesús Zuñiga-Cueva e María Rivera-Chavez: *New trends in pedagogical agents in education*. 2020 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, páginas 923–928, 2020. 7

[23] Zeitlhofer, Ines, Joerg Zumbach e Verena Aigner: *Effects of pedagogical agents on learners' knowledge acquisition and motivation in digital learning environments*. Knowledge, 3(1):53–67, 2023. 7

[24] Cao, Cassie Chen, Zijian Ding, Jionghao Lin e Frank Hopfgartner: *Ai chatbots as multi-role pedagogical agents: Transforming engagement in cs education*. arXiv preprint arXiv:2308.03992, 2023. 7

[25] Morais, Felipe de e Patrícia A. Jaques: *Dinâmica de afetos em um sistema tutor inteligente de matemática no contexto brasileiro: uma análise da transição de emoções acadêmicas*. Revista Brasileira de Informática na Educação, 30:519–541, out. 2022. `https://sol.sbc.org.br/journals/index.php/rbie/article/view/2577`. 7

[26] Siegle, R.F., N.L. Schroeder e H.C Lane: *Twenty-five years of learning with pedagogical agents: History, barriers, and opportunities*. TechTrends, 67:851–864, 2023. 8

[27] Dian Martha, Ati Suci e Harry B. Santoso: *The design and impact of the pedagogical agent: A systematic literature review*. Journal of Educators Online, 16(1), 2019. 8

[28] Paiva, Ana, Isabel Machado e Carlos Martinho: *Enriching pedagogical agents with emotional behaviour—the case of vincent*. AIED, páginas 47–55, 1999. 8

[29] Paiva, Ana e Isabel Machado: *Lifelong training with vincent, a web-based pedagogical agent*. International Journal of Continuing Engineering Education and Life Long Learning, 12(1-4):254–266, 2002. 8

[30] Costa, Evandro de Barros: *Um modelo de ambiente interativo de aprendizagem baseado numa arquitetura multi-agentes*. 1997. 8

[31] Barros, Evandro de, Manoel A. Lopes e Edilson Ferneda: *Mathema: A learning environment based on a multi-agent architecture*. Springer, Berlin, Heidelberg, 2005. 8

[32] Vicari, Rosa Maria, Edilson Pontarolo e Magda Bercht: *Diferentes abordagens de computação afetiva em sistemas multiagentes e sistemas tutores inteligentes*. 2003. 8

[33] Costa, Daniana de e Edilson Pontarolo: *Aspectos da educação ambiental crítica no ensino fundamental por meio de atividades de modelagem matemática*. Revista Brasileira de Estudos Pedagógicos, 100(254), abr. 2019. http://rbep.inep.gov.br/ojs3/index.php/rbep/article/view/3293. 9

[34] *Amazon scraps secret ai recruiting tool that showed bias against women*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK Accessed on 2023-11-28. 10

[35] *Tesla model s driver crashes into a van while on autopilot*. https://electrek.co/2016/05/26/tesla-model-s-crash-autopilot-video/. 10

[36] Cerqueira, José Antonio Siqueira de: *Explorando Ética na elicitação de requisitos em aplicações no contexto de ia*. Universidade de Brasília, 2021. 11

[37] *Lei geral de proteção de dados pessoais, lei n° 13.709/2018*, 2018. https://www.gov.br/governodigital/pt-br/seguranca-e-protecao-de-dados/guias/guia_lgpd.pdf. 11

[38] *Regulamento geral de proteção de dados*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. 11

[39] *Guia de boas práticas lgpd*, 2020. https://www.gov.br/governodigital/pt-br/seguranca-e-protecao-de-dados/guias/guia_lgpd.pdf. 13

[40] Ferber, Jacques, Olivier Gutknecht e Fabien Michel: *From agents to organizations: An organizational view of multi-agent systems*. 2004. 14

[41] Ferber, Jacques e Olivier Gutknecht: *Madkit: A generic multi-agent platform*. Proceedings of the Fourth International Conference on Autonomous Agents., páginas 78–79, 2000. 15, 30

[42] Roussille, Hector, Önder Gürcan e Fabien Michel: *Agr4bs: A generic multi-agent organizational model for blockchain systems*. Big Data and Cognitive Computing, 6(1):1, 2021. 16

[43] Silva, Claiton Custódio da: *Uma ontologia de perfil holístico para estudantes de graduação*. Universidade de Brasília, 2021. 17

[44] Souza, Thiaggo Ferreira Bispo de e Oscar Etcheaverry Barbosa Madureira da Silva: *Implantação de um pod server para ecossistema educacional e sua introdução na educação superior em computação*. Universidade de Brasília, 2023. 18

[45] Duda, João Marcos Schmaltz: *Uma investigação sobre o projeto solid : da prospecção para ecossistema educacional ao desenvolvimento de aplicações como prova de conceito*. Universidade de Brasília, 2023. 18

Anexo A
Matriz de Riscos e Benefícios

Tipo tratamento	Riscos	Mitigações	Benefícios	Descrição	Perguntas Orientadoras
Coleta de dados	Risco da falta de informações sobre as finalidades do proessamento ou falta de engajamento com alunos levar ao não consentimento ou a um consentimento inválido. Risco do não engajamento com alunos levar a baixa utilização da solução.	Engajamento proativo com alunos para colaboração no uso e avaliação da solução.Transparência na apresentação da solução, dos tratamentos de dados realizados e seus fins. Integração com plataforma de diretório de acesso da universidade para autenticar dados de inscrição. Encriptação da comunicação via HTTPS.	Fornecimento do serviço online, cujos benefícios incluem: maior interação com professores e alunos, apoiar na concepção da rede social do CICFriend. Criação de perfil pessoal para interação com controle de acesso.	Avaliar riscos éticos e operacionais da etapa de coleta (operações de coleta, produção e recepção) de dados, especialmente os que tratam dos princípios de: transparência, consentimento informado e agência e responsabilidade do aluno.	Quais os riscos do fornecimento de dados inválidos? Quais os riscos da falta informações levar a um consentimento inválido? Quais os riscos do não engajamento com alunos? Quais os riscos de acesso não autorizado? Quais os riscos da coleção excessiva de dados?
Retenção	Risco da falta de informações sobre as finalidades do proessamento ou falta de engajamento com alunos levar ao não consentimento ou a um consentimento inválido. Risco do não	Controle de acesso lógico, backup de dados do servidor e comunicação clara sobre dados armazenados e seus períodos de retenção.	Fornecimento do serviço online. Criação de uma linha de tempo em forma de publicações.	Avaliar riscos éticos e operacionais na retenção (arquivamento e armazenamento) de dados, especialmente os que tratam dos princípios de: transparência, propriedade e controle de dados e acessibilidade.	Quais os riscos de acesso não autorizado? Quais os riscos de perda de dados? Quais os riscos inerentes do procedimentos para cumprir os prazos de retenção de dados? Quais os dados passíveis de anonimização? Quais dados podem ter tratamentos posteriores? Todos essas possibilidades foram informadas?
Processamento	Risco de enviesamento nas análises levar a bonificação indevida ou a falta dela, risco de extrapolação errônea de dados, risco de falta de consentimento a todo tipo de processamento.	Transparência quanto aos processamentos que serão realizados sobre os dados. Classificação de dados sensíveis e governança de dados.	Fornecimento do serviço online. Possibilidade de receber badges e recompensas por colaborações.	Avaliar riscos éticos e operacionais do processamento de dados, especialmente os que tratam dos princípios de: transparência, propriedade e controle de dados, validade e confiabilidade dos dados, responsabilidade institucional e obrigação de agir, valores culturais, inclusão e consentimento	Quais os riscos de uma classificação/agrupamento enviesado? Quais os riscos de processamento de dados sensíveis? Quais os riscos de reuso não informado? Quais os riscos de tratar os dados mais que o comunicado (transparência)? Quais os riscos de não agir sobre informação? Quais os riscos da extrapolação de dados? Quais os riscos de viés? Quais os riscos da avaliação levar à exclusão? Quais os riscos de modificação não autorizada? Quais os riscos de falha ou erro de processamento?
Compartilhamento	Risco de acesso indevido e riscos de interpretação errada de análises.	Controle de acesso à publicações por listas de acesso, inerentes do Friendica. Procedimentos para comunicação de análises bem definidos.	Fornecimento do serviço online. Rede de amigos podem visualizar e interagir com suas publicações.	Avaliar riscos éticos e operacionais do compartilhamento de dados, especialmente os que tratam dos princípios de: transparência, propriedade e controle de dados e acessibilidade, validade e confiabilidade dos dados, comunicações e consentimento	Quais os riscos de acesso indevido? Quais os riscos de compartilhamento de análises que são estatísticas? Quais os riscos de comunicação indevida como resultado de análises? Quais os riscos de compartilhamento além do consentido?
Eliminação	Risco de perda de dados, risco de reidentificação de dados com tecnologia mais avançada.	Processo de governança de dados bem estabelecido. Anonimização realizada visando a desidentificação completa do aluno.	Possibilidade de eliminação e portabilidade do perfil.	Avaliar riscos éticos e operacionais da eliminação de dados, especialmente os que tratam dos princípios de: transparência, propriedade e controle de dados e consentimento.	Quais os riscos na eliminação indevida de dados? Quais os riscos de perda de dados? Quais os riscos na anonimização de dados antes da eliminação para fins de reuso? Quais os risco de reidentificação de dados anonimizados? Quais os riscos sobre o direito de saber do tratamento de dados ocorrido?