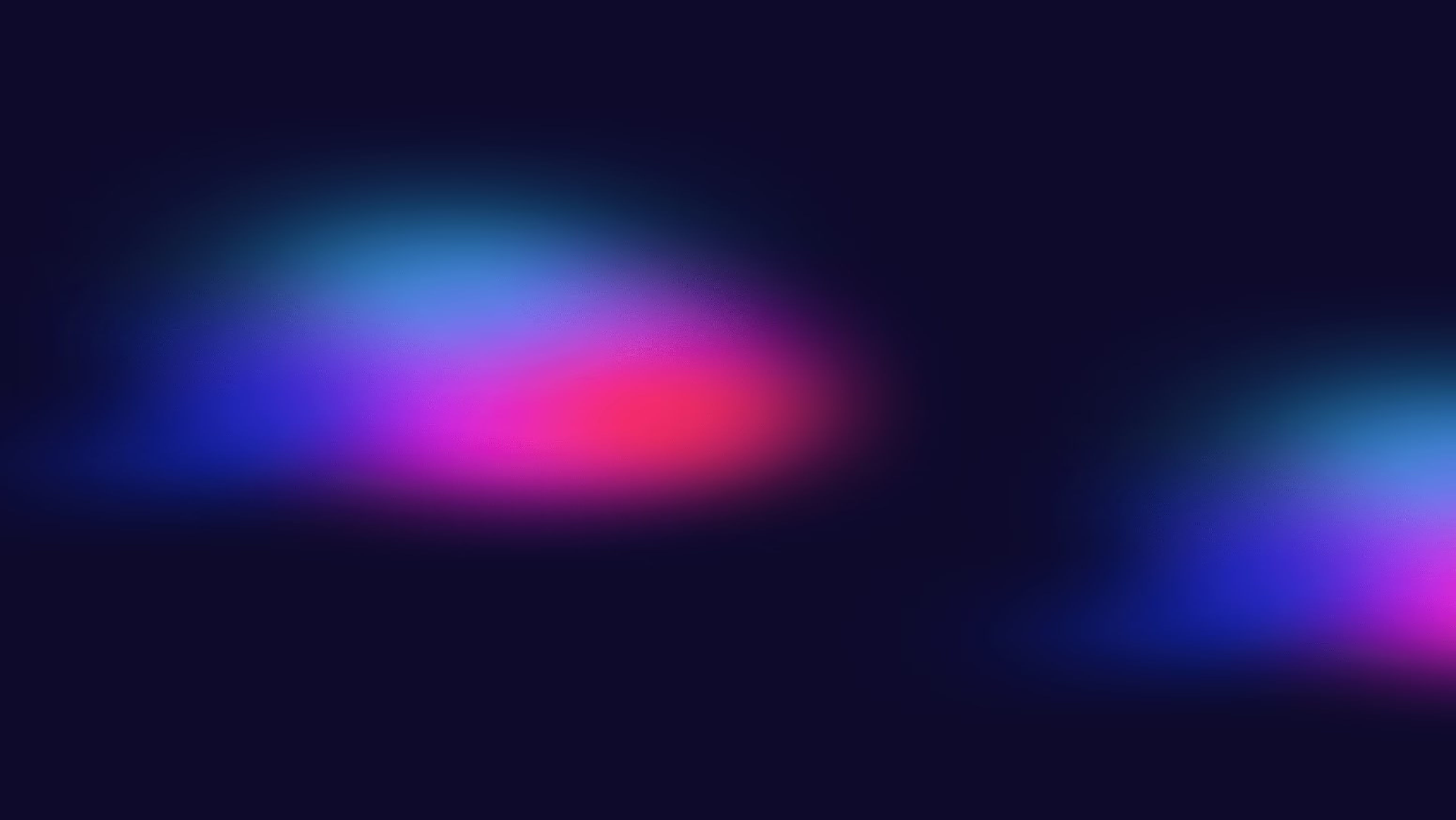
DAB-402 Capstone: AI Chatbot for St. Clair International Team



Ľhis documc"ť ouťli"cs ťhc o"goi"g dc:clopmc"ť or a" AI chaťboť dcsig"cd ťo raciliťaťc i"ťcíacťio"s roí píospccťi:c i"ťcí"aťio"al sťudc"ťs "a:igaťi"g ťhc c"íollmc"ť píoccss ať Sť. Claií Collcgc. Ľhc píojccť ťcam has cmbaíkcd o" a" i""o:aťi:c c"dca:oí, haí"cssi"g ad:a"ccd Naťuíal Ḻa"guagc Píoccssi"g (NḺP) ťcch"iqucs a"d lc:cíagi"g laígc la"guagc modcls ťo cícaťc a" i"ťclligc"ť chaťboť capablc or pío:idi"g accuíaťc a"d íclc:a"ť i"roímaťio" ťo uscís. Ľhíough mcťiculous daťa gaťhcíi"g, pícpíoccssi"g, a"d modcl buildi"g, ťhc ťcam aims ťo dcli:cí a" cxccpťio"al uscí cxpcíic"cc ťhať sťícamli"cs ťhc c"íollmc"ť jouí"cQ.

**bQ Paítk**

# Data Acquisition and Preprocessing

Ľhc píojccť's rou"daťio" lics i" acquiíi"g compíchc"si:c daťa ríom ťwo píimaíQ souíccs: ťhc Sť. Claií I"ťcí"aťio"al Ľcam a"d ťhc collcgc's wcbsiťc. Ľhc ťcam has cmploQcd wcb scíapi"g ťcch"iqucs ťo sQsťcmaťicallQ cxťíacť cíucial i"roímaťio" abouť píogíams, admissio" ícquiícmc"ťs, raculťQ dcťails, a"d couísc orrcíi"gs. Ľhis mcťiculouslQ gaťhcícd daťa has u"dcígo"c íigoíous clca"i"g píoccsscs, i"cludi"g ťhc ícmo:al or cxťía"cous whiťc spaccs, ťo c"suíc co"sisťc"cQ a"d qualiťQ.

## Data Extraction



1

2

3

Wcb scíapi"g ťools, such as «cauťirul Soup i" PQťho", wcíc uťilizcd ťo "a:igaťc ťhc HĽMḺ sťíucťuíc or collcgc wcbsiťcs, sc"d HĽĽP ícqucsťs, paísc HĽMḺ co"ťc"ť, a"d cxťíacť íclc:a"ť daťa. Ľhis píoccss was co"ducťcd cťhicallQ, adhcíi"g ťo ťhc wcbsiťcs' ťcíms or usc a"d ícspo"siblc daťa acquisiťio" píacťiccs.

## Data Cleaning

Arťcí cxťíacťi"g daťa ríom collcgc wcbsiťcs, ťhc ťcam c"cou"ťcícd cxccssi:c whiťc spaccs wiťhi" ťhc scíapcd daťa. Ľo addícss ťhis issuc a"d c"suíc daťa clca"li"css, sQsťcmaťic whiťc spacc ícmo:al píoccsscs wcíc implcmc"ťcd, sig"irica"ťlQ impío:i"g daťa ícadabiliťQ a"d usabiliťQ roí subscquc"ť a"alQsis a"d i"ťcgíaťio" i"ťo ťhc AI chaťboť.

## Tokenization

«oťh chaíacťcí a"d woíd ťokc"izaťio" ťcch"iqucs wcíc cmploQcd ťo a"alQzc ťhc scíapcd daťa. A compíchc"si:c sťudQ íc:calcd ťhať woíd ťokc"izaťio" pío:cd moíc bc"cricial, pío:idi"g a co"ťcxťuallQ íclc:a"ť u"dcísťa"di"g or ťhc daťa íclaťcd ťo collcgc píogíams, admissio" ícquiícmc"ťs, a"d couísc orrcíi"gs, ulťimaťclQ c"ha"ci"g ťhc accuíacQ a"d i"ťcípícťabiliťQ or subscquc"ť NḺP ťasks wiťhi" ťhc AI chaťboť.

# Model Building and Training

Ať ťhc coíc or ťhc AI chaťboť lics a sophisťicaťcd la"guagc modcl ťailoícd spcciricallQ roí Sť. Claií Collcgc. Ľhc ťcam co"ducťcd cxťc"si:c ťíai"i"g usi"g ťhc di:císc daťascť souíccd ríom ťhc collcgc's wcbsiťcs, cmploQi"g boťh chaíacťcí a"d woíd ťokc"izaťio" ťcch"iqucs. UlťimaťclQ, woíd ťokc"izaťio" Qicldcd moíc "ua"ccd i"sighťs i"ťo ťhc collcgc's píogíams, admissio" ícquiícmc"ťs, raculťQ i"roímaťio", a"d couísc orrcíi"gs.

## Custom Language Model

I"iťiallQ, ťhc ťcam cmbaíkcd o" buildi"g a cusťom la"guagc modcl ríom scíaťch, cuíaťi"g a compíchc"si:c daťascť c"compassi"g all íclc:a"ť aspccťs or Sť. Claií Collcgc. Ľhc objccťi:c was ťo imbuc ťhc modcl wiťh a dccp u"dcísťa"di"g or ťhc collcgc's co"ťcxť, c"abli"g iť ťo pío:idc accuíaťc a"d pcíso"alizcd ícspo"scs ťo uscí qucíics.

Mulťihcadcd aťťc"ťio" was uťilizcd i" buildi"g ťhc cusťom la"guagc modcl, allowi"g ťhc modcl ťo rocus o" dirrcíc"ť paíťs or ťhc i"puť scquc"cc simulťa"couslQ, capťuíi"g a íichcí scť or íclaťio"ships a"d paťťcí"s wiťhi" ťhc daťa.

## Pre-trained Models

Rccog"izi"g ťhc "ccd roí c"ha"ccd accuíacQ, ťhc ťcam madc a sťíaťcgic dccisio" ťo ťía"siťio" ríom ťhc cusťom-builť la"guagc modcl ťo lc:cíagi"g píc- ťíai"cd modcls, "amclQ Falco" a"d Ḻlama 2. Ľhcsc modcls, íc"ow"cd roí ťhcií píoricic"cQ i" "aťuíal la"guagc u"dcísťa"di"g, ha:c u"dcígo"c cxťc"si:c ťíai"i"g o" di:císc daťascťs, pío:idi"g a moíc íobusť a"d ícliablc rou"daťio" roí ťhc AI chaťboť's pcíroíma"cc.

# Hyperparameter Tuning and Optimization

Ľo achic:c opťimal pcíroíma"cc a"d accuíacQ, ťhc ťcam has dcdicaťcd sig"irica"ť crroíťs ťo ri"c- ťu"i"g ťhc hQpcípaíamcťcís or ťhc la"guagc modcls cmploQcd. Ľhis iťcíaťi:c píoccss i":ol:cs sQsťcmaťicallQ adjusťi"g :aíious scťťi"gs a"d co"riguíaťio"s, such as lcaí"i"g íaťcs, baťch sizcs, a"d opťimizcí ťQpcs, ťo ri"d ťhc opťimal combi"aťio"s ťhať Qicld ťhc mosť accuíaťc a"d íclc:a"ť ícspo"scs ríom ťhc AI chaťboť.

## Evaluation Metrics



1

2

3

Ľhc ťcam has csťablishcd a íobusť c:aluaťio" ríamcwoík ťo asscss ťhc pcíroíma"cc or ťhc chaťboť. Mcťíics such as pcíplcxiťQ, «ḺEU scoícs, a"d huma" c:aluaťio" aíc cmploQcd ťo qua"ťirQ ťhc qualiťQ or ťhc chaťboť's ícspo"scs a"d idc"ťirQ aícas roí impío:cmc"ť.

## Iterative Refinement

Ľhíough a" iťcíaťi:c píoccss or modcl ťíai"i"g, c:aluaťio", a"d hQpcípaíamcťcí adjusťmc"ť, ťhc ťcam co"ťi"uouslQ ícri"cs ťhc chaťboť's pcíroíma"cc. Ľhis appíoach i":ol:cs sQsťcmaťicallQ ťwcaki"g hQpcípaíamcťcís, ícťíai"i"g ťhc modcl, a"d asscssi"g ťhc impacť o" ťhc c:aluaťio" mcťíics, c"abli"g ťhc ťcam ťo idc"ťirQ ťhc opťimal co"riguíaťio"s.

## Performance Optimization

Ľhc hQpcípaíamcťcí ťu"i"g píoccss is aimcd ať opťimizi"g ťhc chaťboť's pcíroíma"cc, c"suíi"g ťhať iť ca" dcli:cí accuíaťc a"d íclc:a"ť ícspo"scs ťo uscí i"quiíics. «Q ri"c-ťu"i"g ťhc modcl's paíamcťcís, ťhc ťcam aims ťo c"ha"cc ťhc chaťboť's "aťuíal la"guagc u"dcísťa"di"g capabiliťics, c"abli"g iť ťo compíchc"d a"d ícspo"d ťo a widc ía"gc or qucíics crrccťi:clQ.

# Deployment and Integration

Ľhc dcploQmc"ť a"d i"ťcgíaťio" phasc or ťhc píojccť i":ol:cd uťilizi"g Flask, a PQťho" wcb ríamcwoík, a"d ťhc Ľía"sroímcí libíaíQ ťo cícaťc a uscí-ríic"dlQ i"ťcíracc roí ťhc AI chaťboť. I"iťiallQ, ťhc ťcam c"cou"ťcícd challc"gcs i" gc"cíaťi"g ťhc "cccssaíQ "sarc.ťc"soí" rilc ícquiícd roí dcploQmc"ť. Howc:cí, ťhíough pcísc:cía"cc a"d adjusťmc"ťs ťo ťhc modcl's paíamcťcís, ťhc ťcam succcssrullQ gc"cíaťcd ťhc rilc a"d dcploQcd ťhc chaťboť o" a" HĽMḺ plaťroím.

Front-end Development Ľhc ťcam dcdicaťcd crroíťs ťo dc:clopi"g a uscí-ríic"dlQ río"ť-c"d



1

i"ťcíracc, c"suíi"g a scamlcss a"d

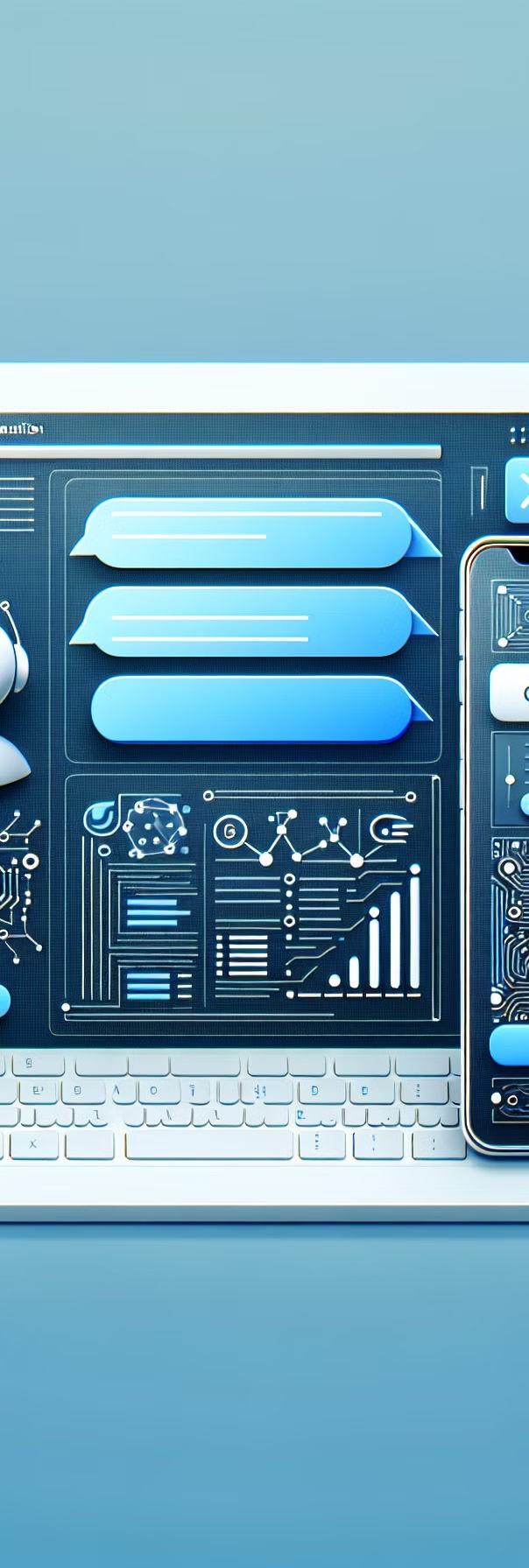
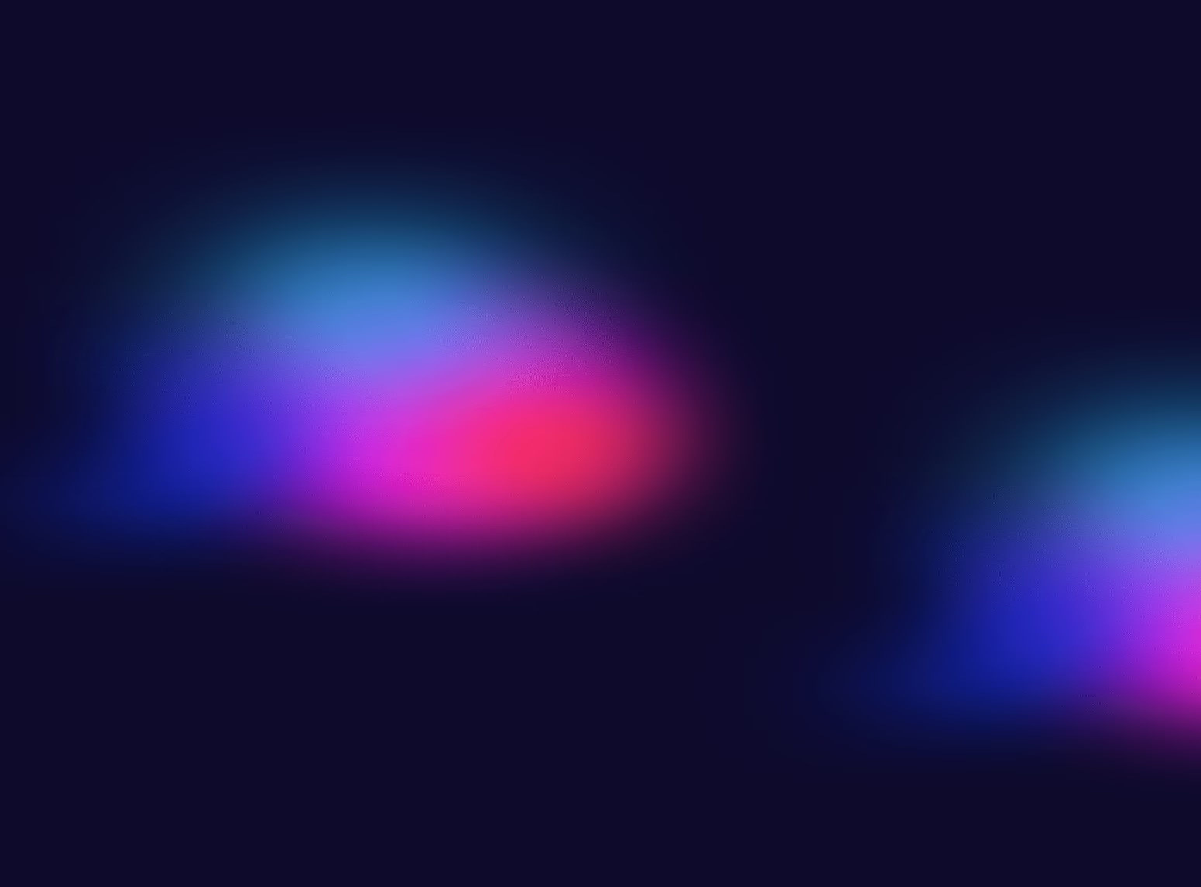
i"ťuiťi:c cxpcíic"cc roí uscís i"ťcíacťi"g wiťh ťhc AI chaťboť. Ľhis i":ol:cd cícaťi"g a :isuallQ appcali"g a"d ícspo"si:c dcsig", i"coípoíaťi"g i"ťuiťi:c chať wi"dows a"d i"puť riclds roí uscís ťo submiť ťhcií qucíics.

## Back-end Integration

O" ťhc back-c"d, ťhc ťcam i"ťcgíaťcd ťhc ťíai"cd la"guagc modcl wiťh ťhc Flask applicaťio", c"abli"g ícal-ťimc commu"icaťio" bcťwcc" ťhc uscí i"ťcíracc a"d ťhc AI chaťboť. Ľhis i":ol:cd implcmc"ťi"g APIs a"d csťablishi"g sccuíc daťa ťía"srcí píoťocols ťo c"suíc crricic"ť a"d sccuíc píoccssi"g or uscí qucíics.



2



## Testing and Optimization



3

Píioí ťo ťhc ri"al dcploQmc"ť, ťhc ťcam co"ducťcd íigoíous ťcsťi"g a"d opťimizaťio" píoccsscs. Ľhis i":ol:cd simulaťi"g :aíious uscí scc"aíios, c:aluaťi"g ťhc chaťboť's ícspo"scs, a"d idc"ťirQi"g poťc"ťial aícas roí impío:cmc"ť. Pcíroíma"cc mcťíics, such as ícspo"sc ťimc a"d accuíacQ, wcíc closclQ mo"iťoícd a"d opťimizcd roí a scamlcss uscí cxpcíic"cc.

# Personalized Course Recommendations

I" a bid ťo c"ha"cc ťhc AI chaťboť's capabiliťics a"d ťía"siťio" iť ríom a mcíc c"íollmc"ť assisťa"ť ťo a compíchc"si:c cducaťio"al ad:isoí, ťhc ťcam aims ťo i"ťcgíaťc pcíso"alizcd couísc íccommc"daťio"s. Ľhis rcaťuíc will lc:cíagc machi"c lcaí"i"g algoíiťhms a"d dcl:c i"ťo hisťoíical c"íollmc"ť paťťcí"s, acadcmic pcíroíma"cc mcťíics, a"d uscí i"ťcíacťio" logs ťo disccí" "ua"ccd pícrcíc"ccs a"d ťailoí couísc íccommc"daťio"s ťo cach sťudc"ť's u"iquc acadcmic jouí"cQ.

## Data Analysis

Ľhc ťcam will co"ducť a" i"- dcpťh a"alQsis or :asť daťascťs c"compassi"g hisťoíical c"íollmc"ť paťťcí"s, acadcmic pcíroíma"cc mcťíics, a"d uscí i"ťcíacťio" logs. Ľhis daťa- díi:c" appíoach will u"co:cí

:aluablc i"sighťs a"d idc"ťirQ paťťcí"s ťhať ca" i"roím pcíso"alizcd couísc íccommc"daťio"s.

## Machine Learning Models

Ḻc:cíagi"g ad:a"ccd machi"c lcaí"i"g algoíiťhms, such as collaboíaťi:c rilťcíi"g, co"ťc"ť- bascd rilťcíi"g, a"d hQbíid íccommc"dcí sQsťcms, ťhc ťcam will dc:clop modcls capablc or pícdicťi"g a"d suggcsťi"g couíscs ťhať alig" wiťh cach sťudc"ť's acadcmic goals, i"ťcícsťs, a"d pcíroíma"cc.

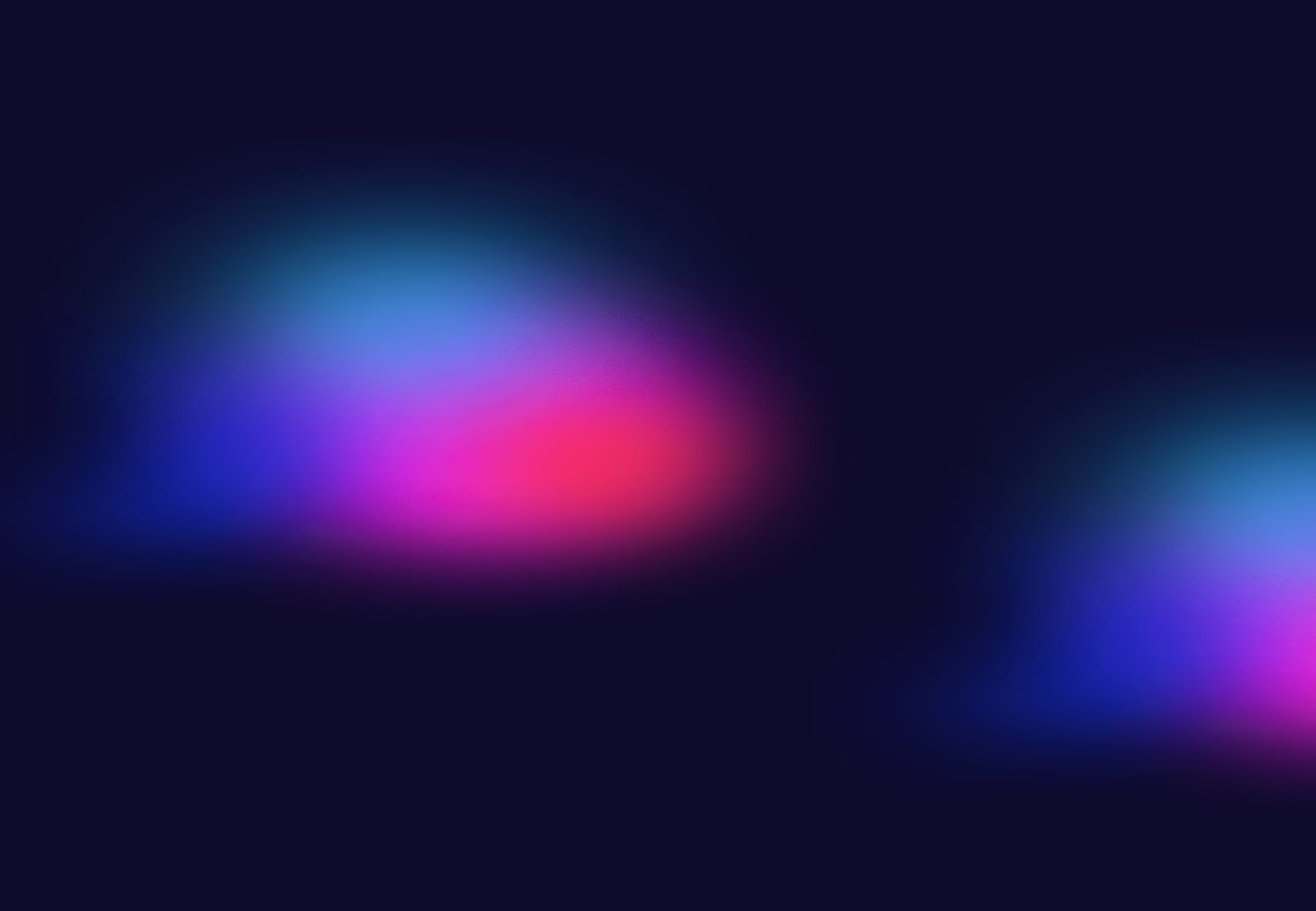
## User Interaction

Ľhc chaťboť will c"gagc i" i"ťcíacťi:c co":císaťio"s wiťh sťudc"ťs, gaťhcíi"g i"roímaťio" abouť ťhcií acadcmic aspiíaťio"s, aícas or i"ťcícsť, a"d pícrcíc"ccs. Ľhis uscí i"puť will bc scamlcsslQ i"ťcgíaťcd wiťh ťhc machi"c lcaí"i"g modcls ťo pío:idc ťailoícd couísc íccommc"daťio"s ťhať opťimizc cach sťudc"ť's cducaťio"al ťíajccťoíQ.

# Ethical Considerations and Responsible Development

As ťhc ťcam cmbaíks o" ťhc dc:clopmc"ť or a" AI-powcícd chaťboť, cťhical co"sidcíaťio"s a"d ícspo"siblc píacťiccs aíc or paíamou"ť impoíťa"cc. Ľhc ťcam is commiťťcd ťo upholdi"g ťhc highcsť sťa"daíds or daťa píi:acQ, sccuíiťQ, a"d ťía"spaíc"cQ, c"suíi"g ťhať ťhc chaťboť's i"ťcíacťio"s wiťh uscís aíc co"ducťcd i" a" cťhical a"d ťíusťwoíťhQ ma""cí.



Data Privacy Ľhc ťcam píioíiťizcs daťa píi:acQ bQ

implcmc"ťi"g íobusť

daťa píoťccťio" mcasuícs, such as c"cíQpťio", acccss co"ťíols, a"d a"o"Qmizaťio" ťcch"iqucs, ťo sarcguaíd uscís' pcíso"al i"roímaťio" a"d mai"ťai" ťhcií ťíusť.

## Fairness and Non- discrimination

Ľhc chaťboť's algoíiťhms a"d modcls aíc dcsig"cd ťo bc u"biascd a"d "o"- discíimi"aťoíQ, c"suíi"g raií a"d cquiťablc ťícaťmc"ť or all uscís ícgaídlcss or ťhcií backgíou"d oí pcíso"al chaíacťcíisťics.

## Transparency and Explainability

Ľhc ťcam sťíi:cs ťo mai"ťai" ťía"spaíc"cQ bQ pío:idi"g clcaí cxpla"aťio"s abouť ťhc chaťboť's ru"cťio"aliťQ, daťa souíccs, a"d dccisio"-maki"g píoccsscs, píomoťi"g ťíusť a"d u"dcísťa"di"g amo"g uscís.

## Security and Robustness

Rigoíous sccuíiťQ mcasuícs aíc implcmc"ťcd ťo píoťccť ťhc chaťboť ríom poťc"ťial ťhícaťs, such as cQbcí-aťťacks, daťa bícachcs, a"d malicious i"puťs, c"suíi"g ťhc sQsťcm's íobusť"css a"d i"ťcgíiťQ.

# User Experience and Continuous Improvement

Ľhc succcss or ťhc AI chaťboť hi"gcs o" dcli:cíi"g a" cxccpťio"al uscí cxpcíic"cc ťhať rosťcís ťíusť, c"gagcmc"ť, a"d saťisracťio" amo"g píospccťi:c sťudc"ťs ať Sť. Claií Collcgc. Ľhc ťcam is commiťťcd ťo co"ťi"uouslQ c"ha"ci"g ťhc chaťboť's capabiliťics ťhíough uscí rccdback, daťa-díi:c" i"sighťs, a"d iťcíaťi:c impío:cmc"ťs.

|  |  |
| --- | --- |
| Uscí Expcíic"cc Facťoí | Appíoach |
| Naťuíal Ḻa"guagc U"dcísťa"di"g | Ḻc:cíagi"g ad:a"ccd NḺP ťcch"iqucs a"d ri"c- ťu"cd la"guagc modcls ťo accuíaťclQ compíchc"d uscí qucíics a"d pío:idc íclc:a"ť ícspo"scs. |
| Co":císaťio"al Flow | Implcmc"ťi"g i"ťclligc"ť dialoguc ma"agcmc"ť sQsťcms ťo mai"ťai" a "aťuíal a"d cohcíc"ť co":císaťio"al rlow, mimicki"g huma"-likc i"ťcíacťio"s. |
| Pcíso"alizaťio" | I"coípoíaťi"g pcíso"alizcd íccommc"daťio"s a"d ťailoícd ícspo"scs bascd o" uscí pícrcíc"ccs, i"ťcícsťs, a"d i"ťcíacťio" hisťoíQ. |
| Co"ťi"uous Ḻcaí"i"g | EmploQi"g machi"c lcaí"i"g algoíiťhms a"d uscí rccdback loops ťo co"ťi"uouslQ impío:c ťhc chaťboť's k"owlcdgc basc a"d ícspo"sc qualiťQ. |

# Project Evaluation and Impact Assessment

As ťhc píojccť píogícsscs, ťhc ťcam íccog"izcs ťhc impoíťa"cc or co"ducťi"g compíchc"si:c c:aluaťio"s a"d impacť asscssmc"ťs ťo mcasuíc ťhc succcss or ťhc AI chaťboť a"d iťs co"ťíibuťio" ťo c"ha"ci"g ťhc c"íollmc"ť cxpcíic"cc ať Sť. Claií Collcgc. Ľhis píoccss i":ol:cs gaťhcíi"g qua"ťiťaťi:c a"d qualiťaťi:c daťa, a"alQzi"g kcQ pcíroíma"cc i"dicaťoís (KPIs), a"d soliciťi"g rccdback ríom sťakcholdcís, i"cludi"g sťudc"ťs, raculťQ, a"d ťhc collcgc admi"isťíaťio".

## Key Performance Indicators (KPIs)



1

Ľhc ťcam will csťablish a scť or KPIs ťo c:aluaťc ťhc chaťboť's pcíroíma"cc, such as uscí c"gagcmc"ť mcťíics, ícspo"sc accuíacQ íaťcs, a"d o:cíall saťisracťio" lc:cls. Ľhcsc KPIs will scí:c as bc"chmaíks roí asscssi"g ťhc chaťboť's crrccťi:c"css a"d idc"ťirQi"g aícas roí impío:cmc"ť.

## User Feedback and Surveys

Gaťhcíi"g diíccť rccdback ríom uscís ťhíough suí:cQs, rocus gíoups, a"d uscí cxpcíic"cc sťudics will pío:idc :aluablc i"sighťs i"ťo ťhc chaťboť's usabiliťQ, crrccťi:c"css, a"d pcícci:cd :aluc. Ľhis rccdback will i"roím ruťuíc c"ha"ccmc"ťs a"d díi:c co"ťi"uous impío:cmc"ť crroíťs.



2

## Stakeholder Engagement

Rcgulaí c"gagcmc"ť wiťh sťakcholdcís, i"cludi"g Sť. Claií Collcgc raculťQ, admi"isťíaťoís, a"d ťhc i"ťcí"aťio"al sťudc"ť commu"iťQ, will c"suíc ťhať ťhc chaťboť alig"s wiťh ťhcií "ccds a"d cxpccťaťio"s. Ľhcií i"puť a"d rccdback will bc cíucial i" shapi"g ťhc chaťboť's dc:clopmc"ť a"d mcasuíi"g iťs o:cíall impacť.



3

# Future Directions and Research Opportunities

Ľhc dc:clopmc"ť or ťhc AI chaťboť roí Sť. Claií Collcgc opc"s up cxciťi"g a:c"ucs roí ruťuíc ícscaích a"d cxploíaťio". As ťhc ricld or "aťuíal la"guagc píoccssi"g a"d gc"cíaťi:c AI co"ťi"ucs ťo c:ol:c íapidlQ, ťhc ťcam is wcll-posiťio"cd ťo capiťalizc o" cmcígi"g ťcch"ologics a"d push ťhc bou"daíics or whať is possiblc i" ťcíms or i"ťclligc"ť co":císaťio"al agc"ťs a"d pcíso"alizcd cducaťio"al cxpcíic"ccs.

## Multimodal AI Integration

I"ťcgíaťi"g mulťimodal AI capabiliťics, such as compuťcí

:isio" a"d spccch íccog"iťio", ca" c"ha"cc ťhc chaťboť's ru"cťio"aliťQ, c"abli"g iť ťo píoccss a"d ícspo"d ťo :aíious roíms or i"puť, such as imagcs,

:idcos, a"d audio íccoídi"gs. Ľhis ca" lcad ťo íichcí a"d moíc immcísi:c cducaťio"al cxpcíic"ccs roí sťudc"ťs.

## Federated Learning and Privacy- Preserving AI

Exploíi"g rcdcíaťcd lcaí"i"g a"d píi:acQ-pícscí:i"g AI ťcch"iqucs ca" c"ablc ťhc chaťboť ťo lcaí" ríom dccc"ťíalizcd daťa souíccs whilc mai"ťai"i"g sťíicť píi:acQ a"d sccuíiťQ sťa"daíds, c"suíi"g ťhať sc"siťi:c uscí daťa is píoťccťcd a"d sccuíc.

## Explainable AI and Interpretability

Ad:a"ci"g ícscaích i"ťo cxplai"ablc AI a"d i"ťcípícťabiliťQ ca" hclp shcd lighť o" ťhc i""cí woíki"gs or ťhc chaťboť's dccisio"-maki"g píoccsscs, píomoťi"g ťía"spaíc"cQ a"d ťíusť amo"g uscís whilc c"abli"g ruíťhcí opťimizaťio" a"d impío:cmc"ť.