

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Tarkvaraarenduse instituut

Sergei Vassiljev 200721IABB

# **GPT-3 kasutamise võimalused ja väljakutsed loomuliku keele töötlemiseks meditsiinis**

Bakalaureusetöö

Juhendaja:

Gunnar Piho

## **Autorideklaratsioon**

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Sergei Vassiljev

20.05.2024

## **Annotatsioon**

### **GPT-3 kasutamise võimalused ja väljakutsed loomuliku keele töötlemiseks meditsiinis**

Kaasaegses meditsiinis seisavad arstid silmitsi vajadusega pühendada olulist aega administratiivsetele ülesannetele, mis vähendab nende töö efektiivsust ja aega, mis veedetakse patsientidega. Uuring on suunatud selle probleemi lahendamisele, kasutades tehisintellekti tehnoloogiat, mida esindab loomuliku keele töötlemise mudel GPT-3, mis võimaldab meditsiinitekstide töötlemise automatiseerimist ja administratiivsete protsesside lihtsustamist.

Peamine eesmärk on uurida GPT-3 potentsiaali ja efektiivsust meditsiinitekstide töötlemisel ja nende teisendamisel lihtsustatud formaati, sarnaselt RDF-iga, et parandada meditsiinilise informatsiooni kättesaadavust ja mõistmist. Eesmärkide saavutamiseks viidi läbi olemasoleva kirjanduse ülevaade AI kasutamisest meditsiinis ning arendati ja testiti GPT-3 põhine süsteemi prototüüp.

Uuringu tulemused näitasid, et GPT-3 kasutamine võib märkimisväärselt lihtsustada meditsiinitekstide töötlemist ja esitamist, vähendada arstidele administratiivsete ülesannete täitmisel kuluvat aega ning seeläbi parandada arstide ja patsientide vahelist suhtlust. See avab uusi võimalusi meditsiiniteenuste parandamiseks ja meditsiinipersonali töö optimeerimiseks.

Uuring rõhutab kaasaegsete AI tehnoloogiate integreerimise tähtsust meditsiinipraktikasse ja pakub suundi edasisteks uuringuteks meditsiinilise informatsiooni töötlemise automatiseerimise valdkonnas. Töö annab panuse tehisintellekti meetodite arendamisse meditsiinis ja pakub praktilisi lahendusi arstide administratiivkoormuse vähendamiseks, mis aitab tõsta meditsiiniteenuste efektiivsust ja kvaliteeti. Rakenduse lähtekood on avaldatud aadressil: <https://github.com/sergeivassiljev/medical-data-rdf>

## **Abstract**

### **The possibilities and challenges of using GPT-3 for natural language processing in medicine.**

In modern medicine, doctors face the necessity to dedicate a significant amount of time to administrative tasks, which reduces the efficiency of their work and the time spent with patients. This study aims to address this issue by utilising artificial intelligence technology, represented by the natural language processing model GPT-3, to automate the processing of medical texts and simplify administrative processes.

The primary goal is to explore the potential and effectiveness of using GPT-3 in processing medical texts and transforming them into a simplified format similar to RDF, to improve the accessibility and understanding of medical information. To achieve these objectives, a review of the existing literature on AI application in medicine was conducted, and a prototype system based on GPT-3 was developed and tested.

The research results showed that the use of GPT-3 could significantly simplify the processing and presentation of medical texts, reduce the time doctors spend on administrative tasks, and thereby improve the quality of interaction between doctors and patients. This opens new prospects for enhancing medical services and optimising the work of medical staff.

The study emphasises the importance of integrating modern AI technologies into medical practice and suggests directions for further research in the field of medical information processing automation. The work contributes to the development of artificial intelligence methods in medicine and offers practical solutions to reduce the administrative burden on doctors, thereby increasing the efficiency and quality of medical services. The source code of the application has been published at the address: <https://github.com/sergeivassiljev/medical-data-rdf>

## Lühendite ja mõistete sõnastik

1. GPT-3 (Generative Pre-trained Transformer 3)	OpenAI poolt arendatud tehisintellekti mudel, mis on üks arenenumaid loomuliku keele töötlemise mudeleid.
2. NLP (loomuliku keele töötlemine)	Tehnoloogia, mis võimaldab arvutitel mõista ja reageerida inimkeelele.
3. RDF (Resource Description Framework)	Veebistandard, mis võimaldab kirjeldada ressursse ja nende suhteid teiste ressurssidega struktureeritud viisil.
4. Kosinussarnasus	Meetod, mida kasutatakse tekstilise sarnasuse mõõtmiseks, arvutades kahe vektori vahelist nurka.
5. Meditsiinilised kirjed	Meditsiinivaldkonnas kasutatavad dokumendid, mis sisaldavad patsientide terviseandmeid.
6. Transformaatori arhitektuur	Neuraalvõrgu arhitektuur, mis on keskne komponent paljudes tänapäevastes NLP mudelites, sealhulgas GPT-sarjas.
7. Eelkoolitus	Protsess, kus mudelit treenitakse suure hulga andmetega enne selle kasutamist spetsiifilistel ülesannetel.
8. Vektoriteks teisendamine	Protsess, kus tekst muudetakse numbriliseks vormiks, et seda saaks töödelda masinõppe mudelitega.
9. Semantiline struktuur	Andmete organiseerimine viisil, mis peegeldab nende tähendust ja suhteid, oluline keeruka informatsiooni, nagu meditsiiniline teave, haldamiseks.
10. Interoperatiivsus	Süsteemide ja organisatsioonide võime vahetada ja kasutada teavet tõhusalt.

11. OpenAI	tehisintellekti uurimisorganisatsioon
12. Web Speech API	on veebitehnoloogia, mis võimaldab veebirakendustel kasutada kõnetuvastuse ja kõnesünteesi funktsioone. Seda API-d pakuvad mitmed veebibrauserid ja see koosneb kahest peamisest osast

## Sisukord

1 Sissejuhatus	7
1.1 Taust	7
1.2 Probleem	8
1.3 Eesmärk	8
2 Metoodika	9
2.1 Objekt	9
2.1.1 Selgitus	10
2.1.2 Oodatav tulemused formaat	10
2.1.3 RDF tripleite organiseerimine	11
2.1.4 Oodatav vastus	11
2.2 Tööriistad	12
2.2.1 Kasutatud programmeerimiskeeled	12
2.2.2 Raamistikud ja teegid	12
2.2.3 NLP mudelid	13
2.2.4 Andmebaas	13
2.2.5 Muud tööriistad	13
2.3 RDF	13
2.3.1 Mis on RDF	13
2.3.2 Miks RDF	15
2.4 Kirjanduse ülevaade	15
2.4.1 Sissejuhatus	15

2.4.2	GPT-3 kasutamise võimalused meditsiinis	16
2.4.3	Meditsiinilise dokumentatsiooni töötlemine ja analüüs	16
2.4.4	Meditsiiniliste aruannete genereerimine	16
2.4.5	Hariduslikud rakendused	16
2.4.6	GPT-3 kasutamise väljakutsed meditsiinis	17
2.4.7	Eetilised ja õiguslikud aspektid	17
2.4.8	Arvutusressursside vajadus	17
2.4.9	Integreerimine olemasolevate süsteemidega	17
2.4.10	Kokkuvõte	18
2.5	GPT arhitektuur ja tööpõhimõte	18
2.6	Tööprotsess	19
3	Töö tulemused	20
3.1	Veebiapp	20
3.2	Häälsisend	23
3.3	Testimine/Analüüs	23
3.3.1	Protsess	24
3.3.2	Praktiline rakendus	24
3.4	API/FLASK	25
3.4.1	Funktsiooni käivitamine veebiliidese kaudu	25
3.4.2	GPT-3 päringu ettevalmistamine	25
3.4.3	Päringu saatmine GPT-3	25
3.4.4	GPT-3 vastuse töötlemine	25
3.4.5	Andmete salvestamine andmebaasi	26
3.4.6	Kasutajale vastuse tagastamine	26
3.5	Prompt (mall)	26
4	Analüüs ja järeldused	27
4.1	Test number 1	28
4.1.1	Toortekst	29
4.1.2	Tulemus	29
4.1.3	Analüüs	30
4.2	Test number 2	33
4.2.1	Toortekst	33
4.2.2	Tulemus	33
4.2.3	Analüüs	35
4.3	Test number 3	36
4.3.1	Toortekst	36
4.3.2	Tulemus	37
4.3.3	Analüüs	38

4.4 Test number 4	39
4.4.1 Toortekst	40
4.4.2 Tulemus	40
4.3.3 Analüüs	42
4.5 Järeldused	43
5 Kokkuvõte	44
Kasutatud allikad	45
Lisad	46
Lisa 1 - “Automatiseeritud testimisskript”	46
Lisa 2 - “Andmete puhastamise skript”	47
Lisa 3 - “TF-IDF ja kosinussarnasuse analüüsi skript tekstiliste andmete jaoks”	47
Lisa 4 - “Tekstidokumentide kosinussarnasuse soojuskaardi visualiseerimise skript”	48
Lisa 5 - “GPT-3.5 Turbo päringu seadistused.”	48

## **Jooniste loetelu**

Joonis 1. RDF skeem ja instantsiandmed	15
Joonis 2. Kontseptuaalne arhitektuur GPT mudelil.[10]	20
Joonis 3. Tulemuste kuvamine	22
Joonis 4. genereeritud RDF	23
Joonis 5. Päringute ajalugu	23
Joonis 6. Meditsiinilise raporti genereerimine	24
Joonis 7. Koosinus Sarnastus 1	32
Joonis 8. Kosinussarnasuse soojuskaart 1	33
Joonis 9. Koosinus Sarnastus 2	36
Joonis 10. Kosinussarnasuse soojuskaart 2	37
Joonis 11. Koosinus Sarnastus 3	39
Joonis 12. Kosinussarnasuse soojuskaart 3	40
Joonis 13. Koosinus Sarnastus 4	43
Joonis 14. Kosinussarnasuse soojuskaart 4	44



# 1 Sissejuhatus

## 1.1 Taust

Kaasaegsed saavutused tehisintellekti ja masinõppe valdkonnas on oluliselt laiendanud loomuliku keele töötlemise (NLP) [2] võimalusi, avades uusi horisonte selle kasutamiseks meditsiinis. GPT-3 (Generative Pre-trained Transformer 3) [1] on üks kõige arenenumaid TI mudeleid NLP jaoks, mille on välja töötanud OpenAI [11], demonstreerides erakordset võimet genereerida teksti, mõista päringuid, tõlkida keeli ja vastata küsimustele. Oma võimsuse ja paindlikkuse tõttu pakub GPT-3 olulisi võimalusi meditsiinisektorile, sealhulgas dokumentatsiooni automatiseerimist, meditsiiniliste kirjade analüüsi, kliinilise otsustusprotsessi toetust ja personaliseeritud nõustamist patsientidele.

## 1.2 Probleem

Viimastel aastakümnetel on meditsiinitööstus kokku puutunud mitmete väljakutsetega, mille seas erilise koha hõivab meditsiinipersonali halduskoormuse suurenemine. Kaasaegsed arstid peavad kulutama olulise osa oma tööajast erinevate infosüsteemidega suhtlemisele, meditsiiniliste andmete töötlemisele ja administratiivsetele ülesannetele. See vähendab aega, mis võiks olla pühendatud otsesele tööle patsientidega.

Minu diplomitöö käigus on üheks peamiseks probleemiks süsteemi genereeritud vastuste potentsiaalne ebatäpsus. GPT-3 täpsuse testimine on kriitilise tähtsusega, kuna mudeli poolt antud tulemused võivad oluliselt mõjutada meditsiiniliste soovitude ja analüüside kvaliteeti ning usaldusväärsust. Eriti oluline on määrata, kui täpselt suudab GPT-3

meditsiinilist teavet töödelda ja tõlgendada, mis on oluline selle tehnoloogia kasutamiseks meditsiinipraktikas.

## 1.3 Eesmärk

Põhieesmärk on testida GPT-3 mudeli rakendatavust meditsiinivaldkondades. Ma püüan välja töötada meetodeid meditsiinitekstidest efektiivseks informatsiooni ekstraheerimiseks ja selle järgnevas struktureerimiseks, mis omakorda peaks vähendama meditsiinitöötajatele dokumentatsiooni töötlemiseks vajalikku aega. See võimaldab neil rohkem aega pühendada otsestele patsiendihoolitsustele, parandades meditsiiniteenuste kvaliteeti ja suhtlustasandit patsientidega. Ma soovin minimeerida tulemuste töötlemise erinevusi ja testida GPT-3 mudeli rakendatavust meditsiinivaldkondades, et hinnata selle efektiivsust ja täpsust selles võtmesektoris.

## 2 Metoodika

### 2.1 Objekt

Minu töö objektiks on patsientide haiguslugude meditsiinilised kirjed, arstide märkmed ja kliiniliste katsete kirjeldused. Peaaegu kompuutertomograafia natiivis ja Kõhu- ja vaagnapiirkond.

RDF [3] vormingu loomiseks tekstist peame kõigepealt välja tooma peamised üksused ja nende suhted, et esitada teavet struktureeritud kujul. Meie teksti RDF-esitusprotsess kasutades Turtle'i süntaksit:

patient:

```

hasClinicalData [
  incidentDate "1 week ago" ;
  lossOfConsciousness "no" ;
  visionImpairmentStartDate "yesterday" ;
  hasSymptom ex:Hematoma ;
  potentialFracture "Fr. orbitae os?" ;
  brainImaging [
    imagingType "native axial plane" ;
    lateralVentriclesStatus "no displacement, not enlarged" ;
    thirdAndFourthVentriclesStatus "on midline, not enlarged" ;
    corticalCerebrospinalFluidSpaces "normal" ;
    brainParenchymaFindings "no lesions" ;
    indicationsOfHemorrhage "none" ;
    indicationsOfAcuteIschemia "none"
  ] ;
  skullFindings "no current findings" ;
  paranasalSinuses "no shadowing"
] ;
summary "No changes due to trauma visible".

```

### 2.1.1 Selgitus

patient: on subjekt, kellel on mitmesugused omadused, mis on seotud tema kliiniliste andmetega (hasClinicalData).

hasClinicalData raames kirjeldavad erinevad omadused kliinilise pildi erinevaid aspekte, sealhulgas teavet kukkumise, teadvuse kaotuse, nägemise halvenemise, hematoomi olemasolu ja muude meditsiinilise läbivaatuse üksikasjade kohta.

brainImaging esindab sisestruktuuri, et kirjeldada aju visualiseerimise tulemusi.

Nii saame käsitsi tõlkida tavalist teksti (meie juhul) RDF-vormingusse.

Vaatame üksikasjalikult läbi ühe näite.

Meil on käsitsi kirjutatud uuring “Peaaju kompuutertomograafia natiivis” kliinilised andmed:

Kukkus nädal tagasi. Teadvust ei kaotanud, silmanägemine halvenenud eilsest. Hematoom. Fr. orbitae os? Uuritud peaaju ja koljupõhimik natiivis aksiaaltasapinnas. III ja IV vatsake on keskjoonel, laienuud ei ole. Külgvatsakesed on dislokatsioonita, laienuud ei ole. Kortikaalsed liikvoriruumid on normaalsed. Koldemuutusi ajukoes ei ole. Verdumisele ega ägedale isheemiale viitavat ei ole Koljuluud aktuaalse leiuta. Ninakõrvalkoopad varjustuseta.

## 2.1.2 Oodatav tulemused formaat

Minu eesmärk on tuvastada tekstis põhiinformatsioon, mida saab esitada RDF tripleitena.

Iga teabe elemendi jaoks anname ülesande määrata subjekt, predikaat ja objekt.

Vormindame need kui <subjekt> <predikaat> <objekt>.

## 2.1.3 RDF tripleite organiseerimine

Pärast kõikide tripleite tuvastamist ja vormindamist, arvutame nende koguarv.

Jagame see koguarv N-ga, et määrata, mitu tripleit ideaalis peaks igas osas olema. Kui ei saa neid võrdselt jagada, siis mõnes osas on üks triple rohkem kui teistes.

Organiseerime tripleid vastavalt N osadeks.

## 2.1.4 Oodatav vastus

Part 1:

- <Patsient> <omabSeisundit> "Vertigo"
- <Patsient> <omabSeisundit> "SAH"
- <Patsient> <omabSeisundit> "Orgaanilise patoloogiata"

Part 2:

- <Patsient> <omabTegevust> "tegeleb spordiga"
- <Patsient> <omabTegevust> "Uuritud peaaju ja koljupõhimik natiivis aksiaaltasapinnas"
- <Patsient> <omabTegevust> "Koljuluud aktuaalse leiuta"

Part 3:

- <Patsient> <omabTestiTulemust> "III ja IV vatsake on keskjoonel, laienuud ei ole"
- <Patsient> <omabTestiTulemust> "Külgvatsakesed on dislokatsioonita, laienuud ei ole"
- <Patsient> <omabTestiTulemust> "Kortikaalsed liikvoriruumid on normaalsed"
- <Patsient> <omabTestiTulemust> "Koldemuutusi ajukoes ei ole"
- <Patsient> <omabTestiTulemust> "Verdumisele ega ägedale isheemiale viitavat ei ole"

Seeläbi me mitte ainult ei loonud teksti RDF formaadis, vaid jagasime selle ka osadeks ja arvutasime, mitu tripletit meil on. Seda näidet võib vaadelda kui täpset vastust, mida me taotlesime.

## 2.2 Tööriistad

### 2.2.1 Kasutatud programmeerimiskeeled

1. Python - peamine programmeerimiskeel, mida kasutati serveri osa, andmetöötluse ja NLP mudelitega integreerimiseks. Python valiti, sest see võimaldab suhelda mudelitega OpenAI API teegi kaudu, millel on klienditeegid mitmes keeles, sealhulgas Pythonis.
2. JavaScript - kasutati interaktiivse kasutajaliidese arendamiseks. JavaScript sobib dünaamiliste veebirakenduste loomiseks ja kliendipoolseks suhtluseks.

### 2.2.2 Raamistikud ja teegid

1. Flask - kergekaaluline veebiraamistik Pythonis, mida kasutati rakenduse serveriosa loomiseks. Flask valiti selle paindlikkuse ja mikroteenuste arhitektuuri toetuse tõttu.
2. React.js - JavaScripti raamistik kasutajaliideste ehitamiseks. React.js kasutatakse reageeriva ja interaktiivse esiosa loomiseks.
3. Bootstrap - veebilehtede kujundamise raamistik, mida kasutatakse kiireks ja kohanduvaks liidese disainiks.

### 2.2.3 NLP mudelid

1. gpt-3.5-turbo-instruct - GPT-3.5-Turbo-Instruct on üks OpenAI tehisintellekti versioone, mis põhineb transformaatori arhitektuuril [6], nagu ka eelmised GPT versioonid. See mudel on GPT-3 täiustatud versioon ja on mõeldud juhiste täitmiseks, st see on optimeeritud kasutaja poolt loomulikus keeles antud käskude paremaks mõistmiseks ja täitmiseks. Mudeli nimetuses "turbo" viitab parandatud jõudlusele ja kiirusele võrreldes GPT-3 baasmudelitega.

2. GPT-3 Davinci, Curie, Babbage - OpenAI GPT-3 mudelite seeria, igaühel neist on erinevad võimsuse tasemed, mida kasutati tutvumiseks.

#### 2.2.4 Andmebaas

1. SQLite3 - kergekaaluline manustatud andmebaasi haldussüsteem, mida kasutati logimiseks ja andmete hoidmiseks. SQLite3 valiti selle integreerimise lihtsuse ja madalate ressursinõudmiste tõttu, kuna mul oli vaja lihtsalt kõigi päringute ajalugu.

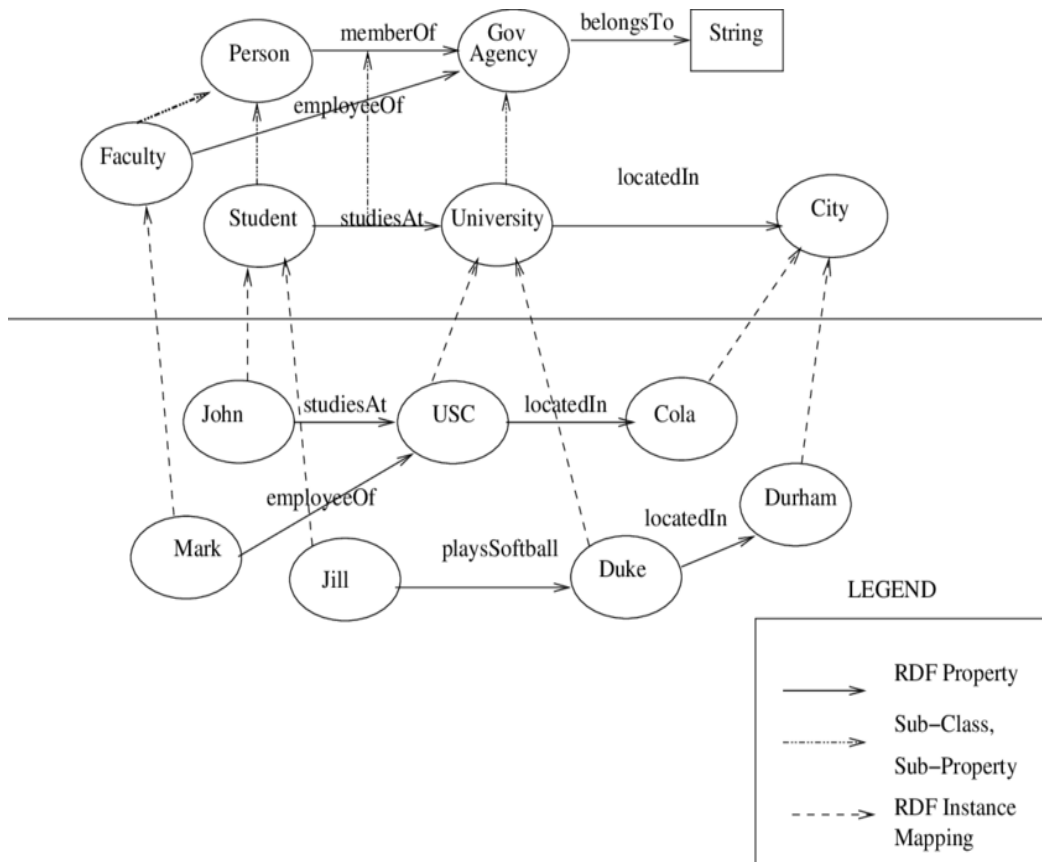
#### 2.2.5 Muud tööriistad

1. SciSpacy - spetsialiseeritud teek meditsiintekstide töötlemiseks, mis põhineb Spacy'l. Kasutatakse meditsiinivaldkonna tekstiandmete analüüsimiseks ja töötlemiseks. Kasutatud tutvumise eesmärgil.

## 2.3 RDF

### 2.3.1 Mis on RDF

RDF ehk Resource Description Framework on andmete kirjeldamise formaat internetis, mis aitab arvutitel "mõista" ja töödelda informatsiooni, muutes selle kättesaadavamaks ja kasulikumaks erinevatele rakendustele. Kujutage seda ette kui universaalset keelt, mis võimaldab erinevatel süsteemidel andmeid vahetada, määratledes objekte (ressursse) ja nende suhteid teineteisega lihtsate väidete kujul. Need väited on üles ehitatud "subjekt-predikaat-objekt" (või "kes-mida-kellega/millega") põhimõttel, muutes informatsiooni struktureerituks ja masinate jaoks kergesti tõlgendatavaks, hõlbustades seeläbi andmete otsimist, analüüsimist ja sidumist võrgus.



Joonis 1. RDF skeem ja instantsiandmed

### 2.3.2 Miks RDF

Ma valisin RDF-i (Resource Description Framework) meditsiinitekstide töötlemiseks mitmel põhjusel:

Semantiline struktuur [9]: RDF võimaldab struktureerida teavet täpse andmete ja nende vaheliste seoste määratlusega, mis on kriitilise tähtsusega keerulise meditsiiniinformatsiooni jaoks.

Kuna RDF on W3C standard, tagab see andmete interoperatiivsuse [10] erinevate süsteemide ja platvormide vahel, parandades seeläbi andmevahetust meditsiinis. RDF-i võime semantiliseks märgistuseks lihtsustab otsinguid ja analüütilisi päringuid, parandades diagnoosimist ja ravivalikuid.



RDF hõlbustab uue teabe ja seoste lisamist, toetades meditsiiniteadmiste pidevat arengut ilma andmebaaside restruktureerimise vajaduseta.

Need tegurid teevad RDF-ist ideaalse valiku efektiivse ja paindliku meditsiiniandmete haldussüsteemi loomiseks, mis aitab parandada patsiendihoolduse kvaliteeti ja teadustööd.

## **2.4 Kirjanduse ülevaade**

### 2.4.1 Sissejuhatus

Loomuliku keele töötlemine (LKT) on üks peamisi tehnoloogiaid, mis toetab meditsiiniliste infosüsteemide arengut. Viimastel aastatel on suurt tähelepanu pälvinud OpenAI mudel GPT-3, millel on kõrge potentsiaal erinevateks meditsiinilisteks rakendusteks. See ülevaade käsitleb GPT-3 kasutamise võimalusi ja väljakutseid meditsiinis, analüüsides praegusi uuringuid ja näiteid rakendustest.

### 2.4.2 GPT-3 kasutamise võimalused meditsiinis

GPT-3 võib kasutada kliiniliste konsultatsioonide automatiseerimiseks ja toetamiseks. Mudel suudab analüüsida sümptomeid, patsiendi haiguslugu ja pakkuda võimalikke diagnoose ja raviplaanid [1]. See on eriti kasulik olukordades, kus juurdepääs meditsiinispetsialistidele on piiratud, näiteks kaugemates piirkondades.

### 2.4.3 Meditsiinilise dokumentatsiooni töötlemine ja analüüs

Meditsiinilise dokumentatsiooni maht kasvab pidevalt ning selle teabe tõhus töötlemine on kriitilise tähtsusega. GPT-3 võib automaatselt välja võtta olulised andmed meditsiinilistest dokumentidest, lihtsustades analüüsi ja suurendades diagnostika täpsust [2]. Näiteks suudab mudel tuvastada olulised sümptomid ja nendega seotud haigused, parandades meditsiiniliste järelduste täpsust.

#### 2.4.4 Meditsiiniliste aruannete genereerimine

Meditsiinispetsialistid kulutavad palju aega aruannete ja dokumentatsiooni koostamisele. GPT-3 võib selle protsessi automatiseerida, genereerides kvaliteetseid meditsiinilisi aruandeid sisestatud andmete põhjal [3]. See vähendab arstide koormust ja soodustab meditsiinilise dokumentatsiooni standardiseerimist.

#### 2.4.5 Hariduslikud rakendused

GPT-3 võib kasutada hariduslikel eesmärkidel, pakkudes üliõpilastele ja spetsialistidele juurdepääsu ulatuslikule meditsiiniliste teadmiste baasile. Mudel suudab vastata keerukatele meditsiinilistele küsimustele, aidates õppijatel paremini mõista erinevaid meditsiini aspekte [4]. See võib olla kasulik ka meditsiinitöötajate pidevas professionaalses arengus.

#### 2.4.6 GPT-3 kasutamise väljakutsed meditsiinis

Üks peamisi väljakutseid GPT-3 kasutamisel meditsiinis on tagada saadud teabe täpsus ja usaldusväärsus. Mudel võib teha vigu või anda ebatäpseid andmeid, mis võib viia tõsiste tagajärgedeni patsientide tervisele [5]. Seetõttu on vaja mudeleid põhjalikult testida ja valideerida enne nende kliinilises praktikas kasutuselevõttu.

#### 2.4.7 Eetilised ja õiguslikud aspektid

GPT-3 kasutamine meditsiinis tõstatab mitmeid eetilisi ja õiguslikke küsimusi. Oluline on arvestada patsientide andmete konfidentsiaalsust ja tagada regulatiivsete nõuete järgimine [6]. Samuti tuleb välja töötada selged juhised LKT-mudelite kasutamiseks, et minimeerida riske ja tagada meditsiinieetika järgimine.

#### 2.4.8 Arvutusressursside vajadus

GPT-3 nõuab märkimisväärseid arvutusressursse nii treenimiseks kui ka kasutamiseks, mis võib piirata selle rakendamist meditsiiniuasutustes, kus tehnilised võimalused on piiratud [7]. See mõjutab ka mudeli kasutuselevõtu ja ekspluateerimise kulusid, mida tuleb arvestada selliste tehnoloogiate kasutamise planeerimisel.

#### 2.4.9 Integreerimine olemasolevate süsteemidega

GPT-3 integreerimine olemasolevate meditsiiniliste infosüsteemidega võib olla keeruline ülesanne. On vaja tagada ühilduvus ja sujuv koostöö erinevate platvormide vahel, et mudelit saaks tõhusalt kliinilises praktikas kasutada [8]. See nõuab nii tehnilisi kui ka organisatsioonilisi jõupingutusi.

#### 2.4.10 Kokkuvõte

GPT-3 omab märkimisväärset potentsiaali erinevate meditsiinipraktikate parandamiseks, sealhulgas kliinilised konsultatsioonid, dokumentatsiooni töötlemine, aruannete genereerimine ja hariduslikud rakendused. Kuid mudeli edukaks kasutuselevõtuks on vaja ületada mitmed väljakutsed, mis on seotud täpsuse, eetika, arvutusressursside ja integreerimisega. Täiendavad uuringud ja vastavate regulatiivsete raamistikude väljatöötamine aitavad kaasa GPT-3 tõhusale ja ohutule kasutamisele meditsiinis.

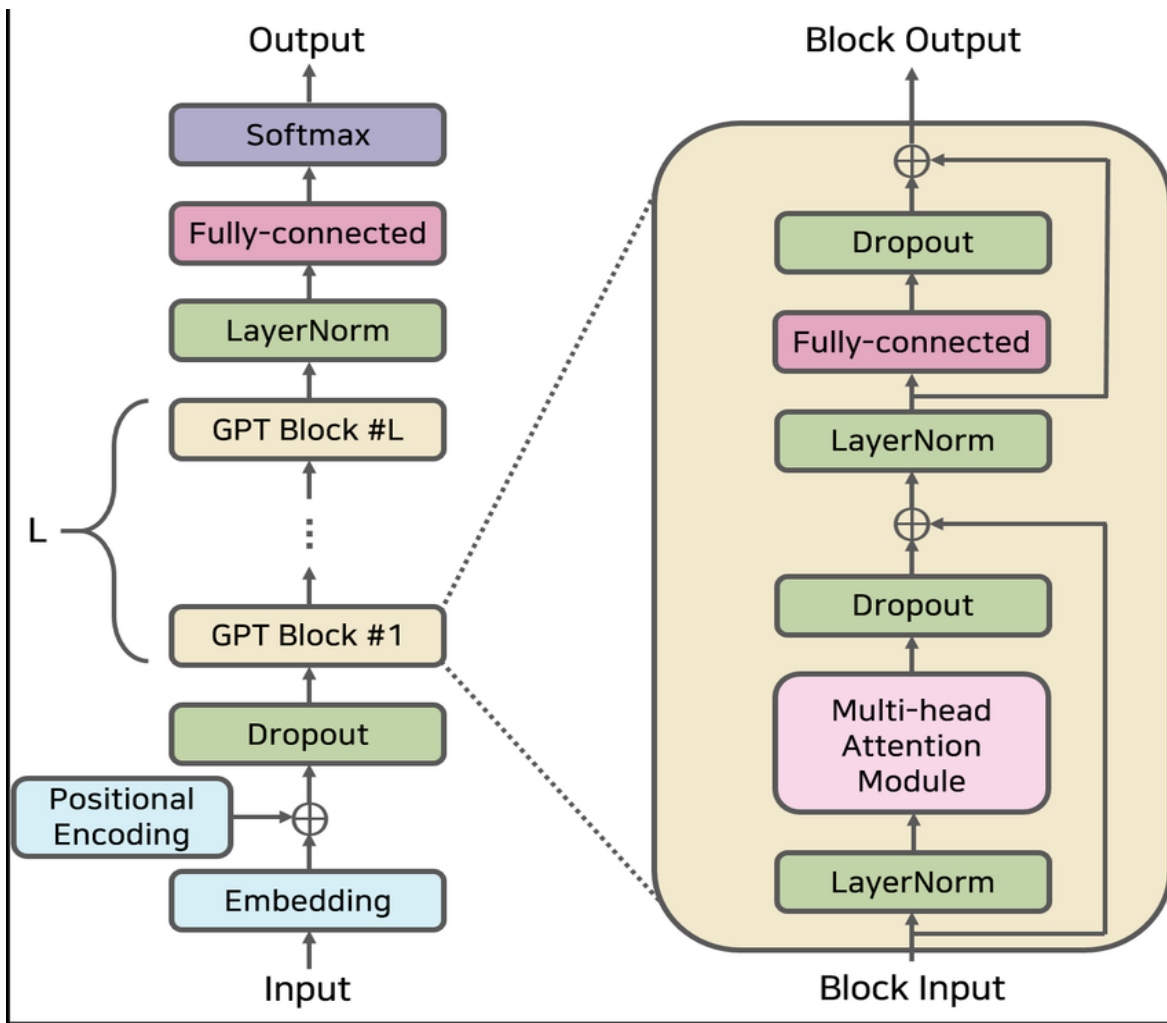
### **2.5 GPT arhitektuur ja tööpõhimõte**

GPT-3 (Generative Pre-trained Transformer 3) arhitektuur on OpenAI poolt välja töötatud edasijõudnud tekstigeneratsiooni mudel. See mudel kasutab transformaatori arhitektuuri, mis esitleti esmakordselt artiklis "Attention is All You Need" autorite Vaswani jt poolt 2017. aastal. GPT-3 on oluliselt laiendanud eelmiste versioonide, sealhulgas GPT ja GPT-2, ulatust ja on tänapäeval üks suurimaid ja võimsamaid loomuliku keele töötlemise mudeleid. [9]

Peamised omadused ja tööpõhimõtted hõlmavad järgmist: Transformerid: GPT aluseks on transformer arhitektuur, mis koosneb self-attention blokkidest ja feed-forward närvivõrkudest. Mudel töötleb teksti, jagades selle tokeniteks (sõnad või sõnaosad), millest igaüht esindatakse vektorkujul.

Self-Attention: Tähelepanu mehhanism võimaldab mudelil määrata, millistele teksti osadele tuleks vastuse genereerimise protsessis rohkem tähelepanu pöörata. See parandab kontekstuaalse mõistmise kvaliteeti ja mudeli võimet töödelda tekstis pikkade sõltuvuste. GPT eeltreenitakse suurel hulgal tekstiandmetel, et saavutada üldine keele mõistmine. Vajadusel saab mudelit peenhäälestada konkreetsete andmetega, et täita kindlaid ülesandeid.

Teksti genereerimine: Vastuse genereerimise protsessis kasutab GPT eelmistel etappidel saadud vektoreid järgmise tokeni ennustamiseks jadas, toetudes järgnevate sõnade võimalike tõenäosuste.



Joonis 2. Kontseptuaalne arhitektuur GPT mudelil.[10]

## 2.6 Tööprotsess

Alguses valmistasin andmed ette. Mulle anti üle tegelikud meditsiinilised kirjed patsientide haiguslugudest, arstide märkmed ja kliiniliste katsete kirjeldused. Tekstid puhastati konfidentsiaalsest informatsioonist ja struktureeriti viisil, mis sobiks mudeli koolitamiseks ja testimiseks.

Seejärel alustasin GPT-3 mudeli seadistamist ja koolitamist: Kohandan GPT-3 mudelit spetsiaalselt meditsiinilise terminoloogia ja konteksti jaoks, kasutades eelpool nimetatud andmeid. See hõlmab lisaseadistusi eelnevalt koolitatud mudelile, et parandada selle võimet mõista ja genereerida meditsiinivaldkonnale iseloomulikke tekste, sealhulgas keerulist terminoloogiat ja lausestruktuure.

Pärast API päringute kaudu saadud vastuste testimist alustasin tekstide formaliseerimise funktsioonide väljatöötamist: Arendasin meetodeid (käsklusi), mis võimaldavad GPT-3 poolt genereeritud tekstid teisendada RDF formaati. See aitab masinal tuvastada ja välja tuua olulist informatsiooni, nagu sümptomid, diagnoosid ja patsiendiandmed.

Seejärel otsustasin alustada lihtsa veebiliidese testimise arendamist: Lõin veebirakenduse React.js ja Flask serveriga, et mugavamalt testida GPT-3 võimalusi.

Testimine ja tulemuste analüüs: pärast NLP seadistamist alustasin süsteemi efektiivsuse kontrollimist, kasutades reaalseid meditsiinijuhtumeid, ja võrdlesin selle täpsust, usaldusväärsust. Testisin umbes 1000-2000 juhtumit ja võrdlesin vastuste täpsust ja sarnasust.

## **3 Töö tulemused**

### **3.1 Veebiapp**

Kasutaja sisestab meditsiinilise teksti tekstivälja kaudu. See tekst saadetakse seejärel serverisse axios.post abil, kui vajutatakse nuppu "Generate RDF".

## Generate RDF from medical text

Enter medical data:

Clinical data: dizziness, nausea. Loss of consciousness 3 days ago, memory disorders. The brain and skull base were examined in the native axial plane. FINDINGS: The lateral ventricles of the brain, III and IV ventricles are of normal shape, size, and position. No focal pathology or changes characteristic of volumetric alteration are evident in the brain tissue. A calcified meningioma on the left frontal side with a diameter of about 1.1 cm. Extracerebral cerebrospinal fluid spaces are of normal width. No intracranial bleeding is detected. Skull bones and skull base are intact. Conclusion: No current findings.



Audio to text:

Generate RDF

Joonis 3. Tulemuste kuvamine

Pärast serverilt vastuse saamist kuvatakse tulemused (genereeritud RDF ja võib-olla tõlgitud RDF) sisestusvormi all olevas kaardis. Kasutaja saab genereeritud ja tõlgitud RDF-i vahel lülituda lülitiga.

Generated RDF

### Generated RDF:

Total RDF Triples: 9 Part 1: - <Patient> <hasCondition> "dizziness" - <Patient> <hasCondition> "nausea" Part 2: - <Patient> <hasCondition> "loss of consciousness" - <Patient> <hasCondition> "memory disorders" Part 3: - <Patient> <hasTest> "examination of brain and skull base" - <Patient> <hasTestResult> "normal shape, size, and position of lateral ventricles" - <Patient> <hasTestResult> "normal shape, size, and position of III and IV ventricles" - <Patient> <hasTestResult> "no focal pathology or changes characteristic of volumetric alteration in brain tissue" - <Patient> <hasTestResult> "calcified meningioma on left frontal side with a diameter of about 1.1 cm" - <Patient> <hasTestResult> "normal width of extracerebral cerebrospinal fluid spaces" - <Patient> <hasTestResult> "no intracranial bleeding detected" - <Patient> <hasTestResult> "intact skull bones and skull base" - <Patient> <hasTestResult> "no current findings"

Get list

Generate Medical Report

Joonis 4. genereeritud RDF

Päringud ja vastused salvestatakse ajalukku, võimaldades kasutajal vaadata eelnevaid tulemusi ja kiiresti nende vahel lülituda.

## History of requests

Clinical data: Upper abdominal pain. Gallstones? B...

Kliinilised andmed: Sapikivid? Kõhukoopa UH uuring...

Kliinilised andmed: Pimesool? Kõhukoopa UH uuring....

Kõhu- ja vaagnapiirkond Kliinilised andmed: Korduv...

Upper abdominal pain. Gallstones? Bile duct dilata...

Joonis 5. Päringute ajalugu



### Generated Medical Report:

Meditsiiniline järelendus: Patsiendil esineb ülemise kõhuvalu ja sapikivid, mis võivad põhjustada sapiteede laienemist. Abdominaalne ultraheliuuring näitas normaalset pankrease ja maksa suurust ning struktuuri. Sapipõie seinad on õhukesed ja luumen on selge, mis viitab sapikivide puudumisele. Sapiteed ei ole laienenud. Põrna ei ole suurenenud ega esine fookuskaebusi. Neerud, kuseteede põis ja eesnääre on normaalsed. Kõhu aord ja iliaalsed veresooned on normaalsed kalibreerid ja vereringe on normaalne. Retroperitoneaalselt pole näha laienenud lümfisõlmi ega vaba vedelikku kõhuõõnes. Eriti tuleks pöörata tähelepanu sapikivide olemasolule, kuna need võivad põhjustada tõsiseid tüsistusi, nagu sapiteede obstruktsioon ja põletik. Patsiendile tuleks soovitada jätkata tervislikku eluviisi ja vajadusel kaaluda sapikivide operatsiooni. Regulaarne jälgimine ja sümptomite kontrollimine on oluline.

Download as PDF

Joonis 6. Meditsiinilise raporti genereerimine

Kasutaja saab genereerida meditsiinilise aruande järelendusena ja samuti märgime, millele tuleks erilist tähelepanu pöörata. (Seda funktsionaalsust saab parandada)

renderRDFParts: Funktsioon jagab RDF-teksti osadeks, lähtudes tekstis olevatest jaotistest ja alajaotistest. Iga osa kuvatakse loendina, mis võimaldab hõlpsasti orienteeruda RDF-struktuuris.

## 3.2 Häälisisend

Katseid tehes otsustasin, et oleks suurepärase idee, kui kasutaja saaks teksti sisestada mitte ainult klaviatuuri abil, vaid ka häälega. Otsustasin proovida kõige lihtsamat meetodit. Kasutasin Web Speech API-d ja pärast teksti saamist hääle abil kontrollisin seda punktuatsiooni-, grammatika- ja trükivigade osas.

## 3.3 Testimine/Analüüs

### 3.3.1 Protsess

Esitasin sama meditsiinilise teksti GPT-3 mudelile mitusada korda. See võimaldab koguda piisavalt andmeid statistiliseks analüüsiks.

Kõik saadud vastused salvestatakse edasiseks analüüsiks. See hõlmab ka toor-RDF tekste.

Tekstide teisendamine vektoriteks [8]: Kosinussarnasuse [4] kasutamiseks teisendatakse vastuste tekstid numbrilisteks vektoriteks.

Iga vastuste paari vahel arvutatakse kosinussarnasus, mis võimaldab kindlaks teha, kui sarnased on vastused üksteisele.

Tulemuste tõlgendamine:

Kui vastuste vahelise kosinussarnasuse keskmine väärtus on kõrge, näitab see, et mudel pakub stabiilseid ja järjepidevaid vastuseid samale päringule.

Madalad kosinussarnasuse väärtused võivad viidata sellele, et mudel genereerib mitmekesiseid vastuseid, mis võivad olla seotud mudeli päringu tõlgendamise varieeruvusega.

Kosinussarnasuse väärtuste standardhälve näitab, kui erinevad on vastused omavahel. Suur standardhälve võib viidata mudeli ebastabiilsusele.

### 3.3.2 Praktiline rakendus

Analüüsi tulemusi saab kasutada mudeli parameetrite täiustamiseks, et tagada stabiilsemad vastused.

Vastuste erinevuste mõistmine aitab tuvastada potentsiaalseid vigu või mudeli nõrkusi, mis vajavad korrigeerimist.

Tulemusi saab kasutada ka mudeli usaldusväärsuse hindamiseks tundlike andmete, nagu meditsiiniline informatsioon, töötlemise ülesannetes.

## 3.4 API/FLASK

### 3.4.1 Funktsiooni käivitamine veebiliidese kaudu

Funktsioon `generate_rdf` aktiveeritakse, saates POST-päringu URL-ile `/generate_rdf`. See meetod ootab JSON-andmeid, mis sisaldavad meditsiinilist teksti.

Sisendandmete lugemine:

Päringust saadud andmed ekstraheeritakse JSON-vormingust. Muutujasse `text` salvestatakse tekst, mis on saadud JSON-objekti väljalt `'text'`.

### 3.4.2 GPT-3 päringu ettevalmistamine

Kõigepealt loeb programm päringu malli failist `rdf_prompt_template.txt`. See mall sisaldab eelnevalt määratletud päringu struktuuri, mis tuleb saata GPT-3-le.

Seejärel asetatakse päringust saadud meditsiiniline tekst sellesse malli, moodustades lõpliku päringu teksti (prompt).

### 3.4.3 Päringu saatmine GPT-3

Koostatud päring saadetakse GPT-3 mudelile (konkreetsemalt `gpt-3.5-turbo-instruct`) OpenAI API kaudu. Päringus määratakse parameetrid vastuse genereerimiseks, nagu temperatuur (`temperature`), maksimaalne tokenite arv (`max_tokens`) ning parameetrid, mis kontrollivad vastuste unikaalsust ja sagedust (`top_p`, `frequency_penalty`, `presence_penalty`). Vaadake Lisa 5 projekti seadistuste jaoks.

### 3.4.4 GPT-3 vastuse töötlemine

GPT-3 vastus sisaldab genereeritud teksti, mis ekstraheeritakse ja töödeldakse edasi. See tekst esitab informatsiooni RDF-vormingus, mis on struktureeritud spetsiaalselt järgneva kasutamiseks meditsiinilistes ja teadusuuringute süsteemides.

### 3.4.5 Andmete salvestamine andmebaasi

Genereeritud RDF-tekst salvestatakse seejärel andmebaasi arhiveerimiseks ja edasiseks kasutamiseks. See võimaldab süsteemil jälgida ja analüüsida teenuse kasutamist ja tulemusi.

### 3.4.6 Kasutajale vastuse tagastamine

Lõpetuseks tagastab programm RDF-teksti JSON-vormingus vastusena algele POST-päringule.

## 3.5 Prompt (mall)

Mallis on selgelt öeldud, et GPT-3 peab meditsiinilise teksti konverteerima RDF-ks, mis esitab andmeid triplitena. See hõlmab teksti analüüsimist olulise teabe tuvastamiseks ja selle struktureerimist vastavas vormingus.

#### 1. Sisendandmed

Input Medical Text: Sellesse muutujasse sisestatakse töödeldav meditsiiniline tekst. See tekst lisatakse päringusse, võimaldades GPT-3 mudelil mõista konteksti ja sisu töötlemiseks.

#### Töötlemisjuhised

Olulise teabe tuvastamine: GPT-3 analüüsib teksti teabe osas, mida saab esitada RDF-triplite kujul.

Iga leitud teabe jaoks peab GPT-3 määrama subjekti, predikaadi ja objekti.

2. Teave tuleb esitada triplitena <subject> <predicate> <object>.

#### RDF-triplite korraldus

3. Pärast kõigi triplite tuvastamist loeb GPT-3 nende koguarvu.

4. Triplid korraldatakse N osaks vastavalt juhistele, arvestades, et mõned osad võivad sisaldada ühte triplit rohkem, kui jaotus ei ole ühtlane.

#### 5. Väljundi vorming

Väljastatakse tuvastatud RDF-triplite koguarv ja nende jaotus osade kaupa vastavalt antud juhistele.

## 6. Väljundi näide

Mall sisaldab ka näidet selle kohta, kuidas vastus peaks välja nägema. See aitab mudelil paremini mõista vajaliku vastuse struktuuri ja vormingut, tagades täpsema ülesande täitmise.

Selline lähenemine võimaldab GPT-3 mudelil efektiivselt ja täpselt töötada meditsiiniliste tekstidega, muutes need struktureeritud ja kergesti töödeldavaks vorminguks.

## 4 Analüüs ja järeldused

Ma kirjutasin Pythoni skripti, mida kasutatakse testimiseks ja andmete kogumiseks. Programm suhtleb Flask-serveriga, et automaatselt genereerida vastuseid meditsiinilistele päringutele RDF-vormingus. Skripti iga komponendi töö üksikasjalik kirjeldus:

### 1. Mallifaili lugemine

Funktsioon `read_query_template` võtab faili teekonna (`file_path`) ja loeb faili sisu. Fail sisaldab päringu teksti malli, mida kasutatakse meditsiiniliste tekstide kujundamiseks, mis saadetakse Flask-serverile. Funktsioon tagastab faili sisu stringina.

### 2. Päringu saatmine Flask-serverile

Funktsioon `send_request_to_flask` võtab kaks parameetrit: `endpoint` (Flask-serveri URL) ja `text` (päringu tekst). Funktsioon saadab POST-päringu määratud `endpoint`-ile koos meditsiinilise tekstiga JSON-vormingus. Kui server vastab veatult (vastuse staatus 200), tagastab funktsioon tulemuse JSON-vormingus. Veakorral tagastatakse `None`.

### 3. Päringute teostamise ja tulemuste salvestamise põhifunktsioon

Funktsioon ``perform_requests_and_save_to_excel`` on skripti põhifunktsioon. See võtab vastu serveri URL-i (``endpoint``), teksti malli (``text_template``) ja päringute arvu (``num_requests``). Selle funktsiooni raames:

- Initsialiseeritakse nimekiri ``results`` päringute tulemuste hoidmiseks.
- For-tsükliga korratakse päringute saatmise protsessi serverile, korduste arv on võrdne ``num_requests``-iga.
- Igal iteratsioonil vormindatakse meditsiiniline tekst malli ``text_template`` abil, asendades ``"PROMT"`` aktuaalse meditsiinilise tekstiga, ja tulemus lisatakse nimekirja ``results``.
- Pärast kõigi päringute sooritamist salvestatakse tulemused Exceli faili abil pandas teeki kasutades.

See skript (vaata Lisa 1) võimaldab automaatiseerida testimisprotsessi käesoleva töö raames, tagades andmete tõhusa kogumise ja organiseerimise edasiseks analüüsiks ja GPT-3 mudeli hindamiseks meditsiiniliste tekstide rakendamisel RDF-vormingus.

Edasi laadib programm andmed Exceli failist, puhastab need erimärkidest ja salvestab tulemuse uude Exceli faili. (Lisa 2)

Edasi tuleb skript, mis on mõeldud tekstiliste andmete analüüsimiseks, kasutades TF-IDF (term frequency-inverse document frequency) meetodit ja kosinussarnasust tekstide sarnasuse määra hindamiseks. (Lisa 3)

Edasi tegin skripti tekstidokumentide kosinussarnasuse visualiseerimiseks soojuskaardina (heatmap), mis võimaldab selgelt hinnata dokumentide omavahelist lähedust. (Lisa 4)

## **4.1 Test number 1**

Päringute arv: (500)

Tokenite arv: (~970000)

#### 4.1.1 Toortekst

Kliinilised andmed: pearinglus, iiveldus. Teadvusekaotus 3 päeva tagasi, mäluhäired. Aju ja kolju põhi uuriti natiivses aksiaaltasapinnas. LEIUD: Aju külgvatsakesed, III ja IV vatsakesed on normaalse kuju, suuruse ja asendiga. Aju koes ei ole nähtavaid koldepatoloogiaid ega ruumala muutustele iseloomulikke muutusi. Vasakul frontaalsel küljel on kaltsifitseerunud meningeoom, läbimõõduga umbes 1,1 cm. Ekstratserebraalsed tserebrospinaalvedeliku ruumid on normaalse laiusega. Koljusisest verejooksu ei tuvastata. Kolju luud ja kolju põhi on terved. Kokkuvõte: Hetkel leidudeta.

#### 4.1.2 Tulemus

Järjestatud

Part 1:

<Patsient> <onSeisund> "pearinglus"

<Patsient> <onSeisund> "iiveldus"

Part 2:

<Patsient> <onSümptom> "teadvusekaotus"

<Patsient> <onSümptom> "mäluhäired"

Part 3:

<Aju> <onUuritud> <Kolju põhi>

<Kolju põhi> <onKuju> "normaalne"

<Kolju põhi> <onSuurus> "normaalne"

<Kolju põhi> <onAsend> "normaalne"

Part 4:

<Ventriklid> <onKuju> "normaalne"

<Ventriklid> <onSuurus> "normaalne"

<Ventriklid> <onAsend> "normaalne"

Part 5:

<Aju kude> <onPatoloogia> "puudub"

<Aju kude> <onMuutused> "puudub"

Part 6:

<Meningioom> <onAsukoht> "vasak frontaalne külg"

<Meningioom> <onLäbimõõt> "1,1 cm"

<Meningioom> <onKaltsifikatsioon> "olemas"

Part 7:

<Tserebrospinaalvedeliku ruumid> <onLaius> "normaalne"

Part 8:

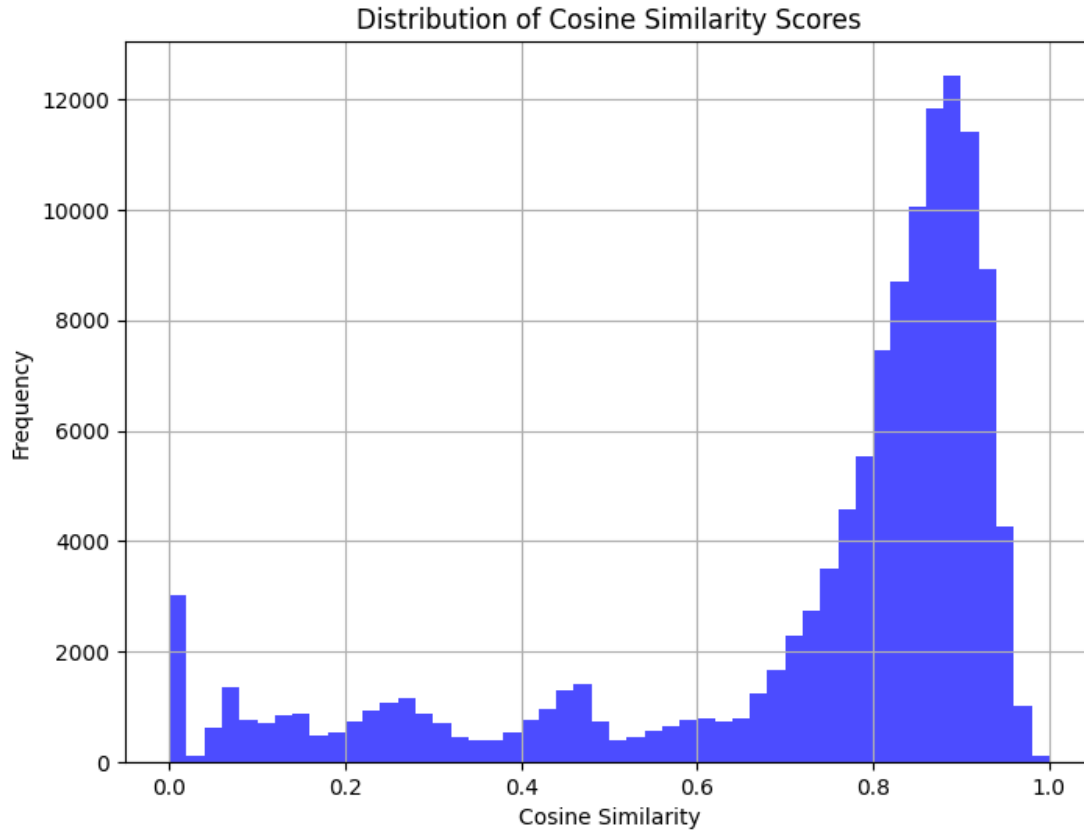
<Kolju luud> <onTerviklikkus> "terved"

Part 9:

<Kokkuvõte> <onLeidud> "puuduvad"

#### 4.1.3 Analüüs





Joonis 7. Koosinus Sarnastus 1

Koosinus Sarnaste analüüs esimese vastuse ja teiste vahel näitas järgmisi tulemusi:

Keskmine sarnasus: 0.794

Standardhälve: 0.194

Maksimaalne väärtus: 1.0

Mediaan (50%): 0.863



Joonis 8. Kosinussarnasuse soojuskaart 1

Soojemad värvid tähistavad kõrgemaid sarnasusi, samas kui jahedamad värvid näitavad madalamat sarnasust.

Jaotuse graafikult on näha, et enamik sarnasuse väärtusi on koondunud kõrgemate väärtuste ümber, mis näitab esimese vastuse ja teiste vahel üsna kõrget sarnasust.

Kosinus Sarnasus on mõõt, mis näitab sarnasust kahe vektori vahel mitmemõõtmelises ruumis, mis meie juhul võib olla kasutusel meditsiiniliste tekstide sarnasuse hindamiseks pärast nende teisendamist RDF formaati. Põhinedes saadud statistilistele andmetele:

Keskmine sarnasus on 0.794. See on üsna kõrge näitaja, mis näitab, et teisendatud tekstidel on oluline sarnasuse aste.

NLP ja AI kontekstis peetakse kosinussarnasust 0.794 keskmise näitajana üsna kõrgeks. See tähendab, et teisendatud tekstid näitavad olulist sarnasust nende vektoriaalsetes esitustes. Kõrge kosinussarnasus viitab minu meetodi efektiivsusele sarnaste tekstide tuvastamisel ja grupeerimisel.

## 4.2 Test number 2

Päringute arv: (100)

Tokenite arv: (~107200)

### 4.2.1 Toortekst

Kliinilised andmed: Insult? Leitud põrandalt, elab üksi. Tehti aju natiivne KT-skaneerimine aksiaaltasapinnas. LEIUD: Aju külgvatsakesed ning kolmas ja neljas vatsake on normaalse kuju ja asetusega, laienuvad. Aju koes ei täheldata koldepatoloogiat ega ruumala muutustele iseloomulikke muutusi. Ekstratserebraalsed tserebrospinaalvedeliku ruumid on laienuvad. Koljusisest hemorraagiat ei tuvastata. Kolju luud ja kolju põhi on terved. KOKKUVÕTE: Tserebrospinaalvedeliku ruumide laienemine - vanusega seotud muutus.

### 4.2.2 Tulemus

Kokku RDF Triples: 10

<Patsient> <onSeisund> "Insult"

<Patsient> <onSeisund> "Tserebrospinaalvedeliku ruumide laienemine - vanusega seotud muutus"

Part 2:

<Patsient> <onTest> "Aju KT-skaneerimine"

<Patsient> <onTest> "Aksiaaltasapind"

Part 3:

<Patsient> <onTestiTulemus> "Normaalne kuju ja asetus"

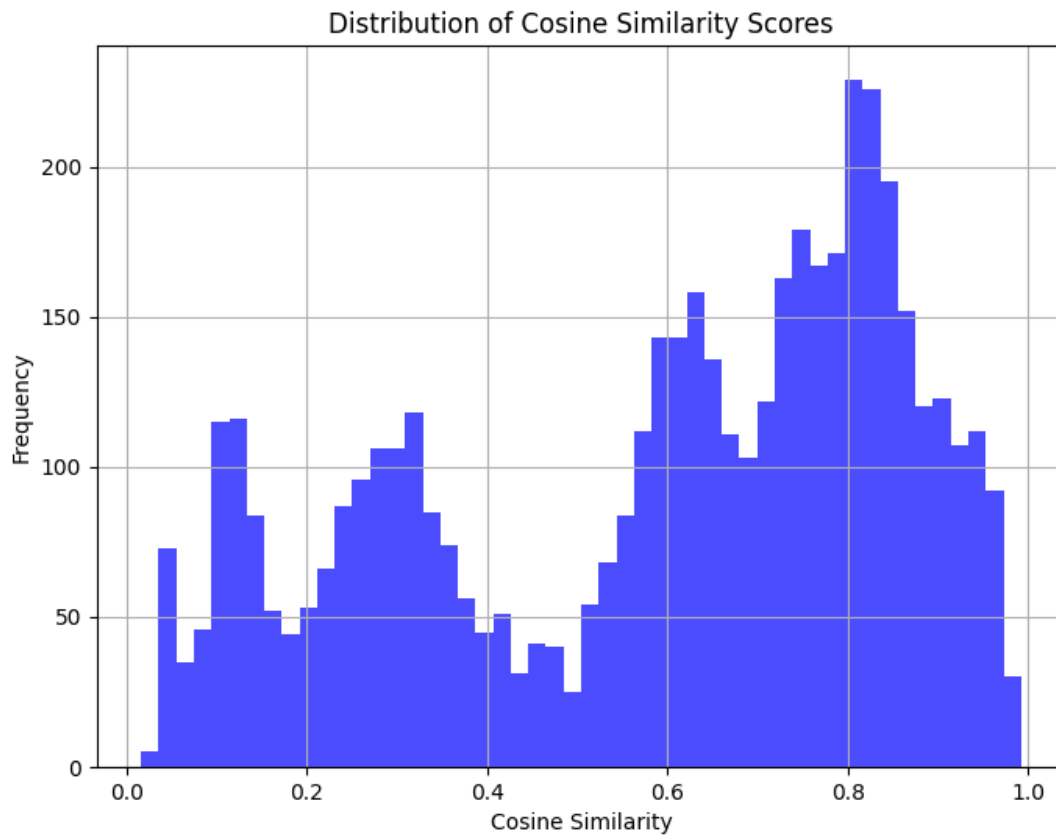
<Patsient> <onTestiTulemus> "Aju koes ei täheldata koldepatoloogiat ega ruumala muutustele iseloomulikke muutusi"

<Patsient> <onTestiTulemus> "Ekstratserebraalsed tserebrospinaalvedeliku ruumid on laienenud"

<Patsient> <onTestiTulemus> "Koljusisest hemorraagiat ei tuvastata"

<Patsient> <onTestiTulemus> "Kolju luud ja kolju põhi on terved"

### 4.2.3 Analüüs



Joonis 9. Koosinus Sarnastus 2

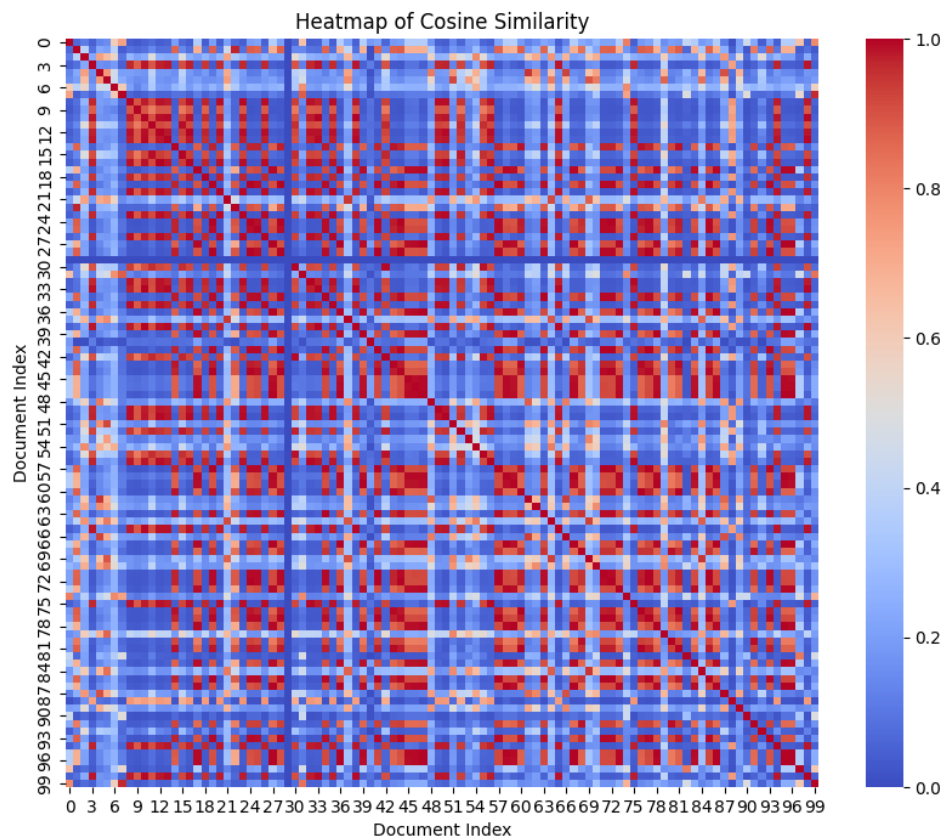
Keskmine sarnasuse väärtus: 0.354

Standardhälve: 0.362

Minimaalne sarnasuse väärtus: 0.124

Mediaan sarnasuse väärtus: 0.179

Maksimaalne sarnasuse väärtus: 1.000



Joonis 10. Kosinussarnasuse soojuskaart 2

### 4.3 Test number 3

Päringute arv: (100)

Tokenite arv: (~84000)

#### 4.3.1 Toortekst

Kõhu- ja vaagnapiirkonna kliinilised andmed: Kukkus kelgutades. LEIUD: Pankreas on normaalse suurusega ja vanusele vastava struktuuriga, ruumala muutusi ei täheldata. Maks on normaalse suurusega ja ühtlase struktuuriga, koldelisi kahjustusi ei tuvastata.

Sapipõie seinad on õhukesed, valendik on selge. Sapijuhad ei ole laiunud. Põrn ei ole suurenenud, koldelisi kahjustusi ei leita. Neerude suurus, asend ja struktuur on normaalsed. Neerudes turset ei esine. Kusepõis on minimaalselt täidetud, sisu on kajavaba. Nähtavas osas on kõhuaort ja niudearterid normaalse läbimõõduga, verevool on normaalne. Retroperitoneaalselt suurenenud lümfisõlmi ei nähtu. Vaba vedelikku kõhuõõnes ei leidu.

#### 4.3.2 Tulemus

Total RDF Triples: 14

Part 1:

<Patsient> <onSeisund> "Kukkus kelgutades."

Part 2:

<Pankreas> <onSuurus> "normaalne suurus"

<Pankreas> <onStruktuur> "vanusele vastav struktuur"

<Maks> <onSuurus> "normaalne suurus"

<Maks> <onStruktuur> "ühtlane struktuur"

<Sapipõis> <onSeinaPaksus> "õhuke"

<Sapipõis> <onLumen> "selge"

<Sapijuhad> <onLaiendatud> "mitte laiunud"

<Põrn> <onSuurenenud> "mitte suurenenud"

<Neerud> <onSuurus> "normaalne suurus"

<Neerud> <onAsend> "normaalne asend"

<Neerud> <onStruktuur> "normaalne struktuur"

<Kusepõis> <onTäidetud> "minimaalselt täidetud"

<Kusepõis> <onSisu> "kajavaba"

Part 3:

<Kõhuaort> <onLäbimõõt> "normaalne läbimõõt"

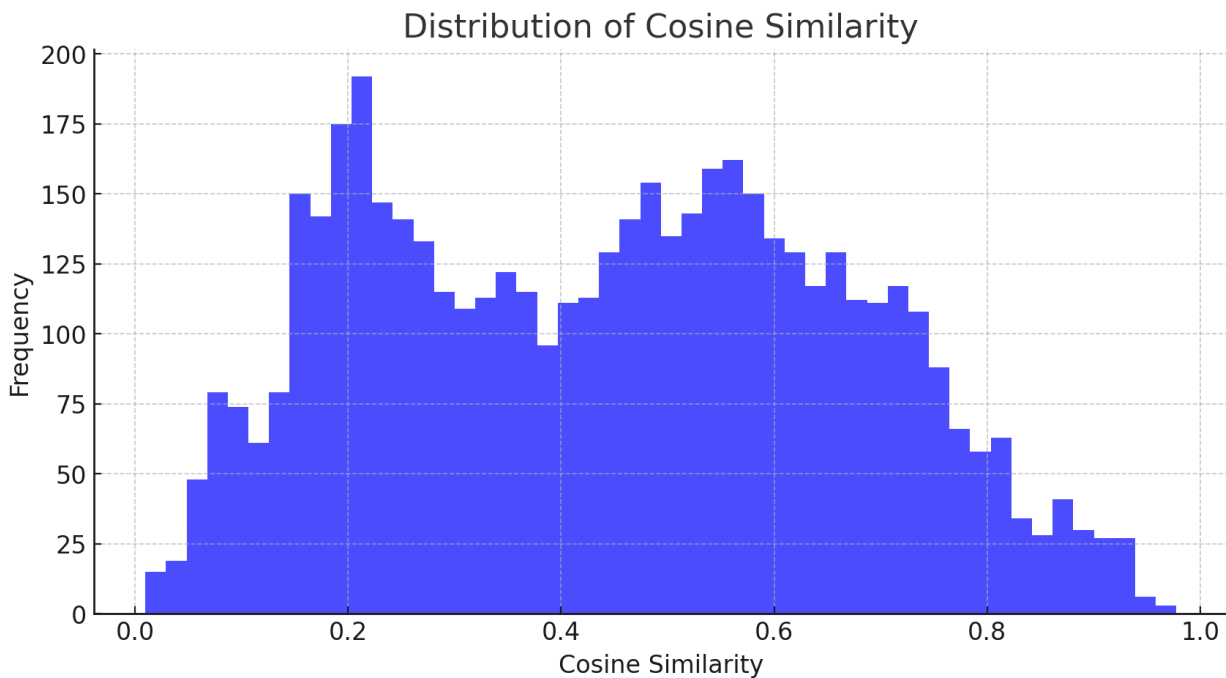
<Niudearterid> <onLäbimõõt> "normaalne läbimõõt"

<Verevool> <onNormaalne> "normaalne"

<Lümfisõlmed> <onSuurenenud> "mitte suurenenud"

<Kõhuõõs> <onVedelik> "vaba vedelik puudub"

### 4.3.3 Analüüs



Joonis 11. Koosinus Sarnastus 3

Keskmine sarnasus: 0.447

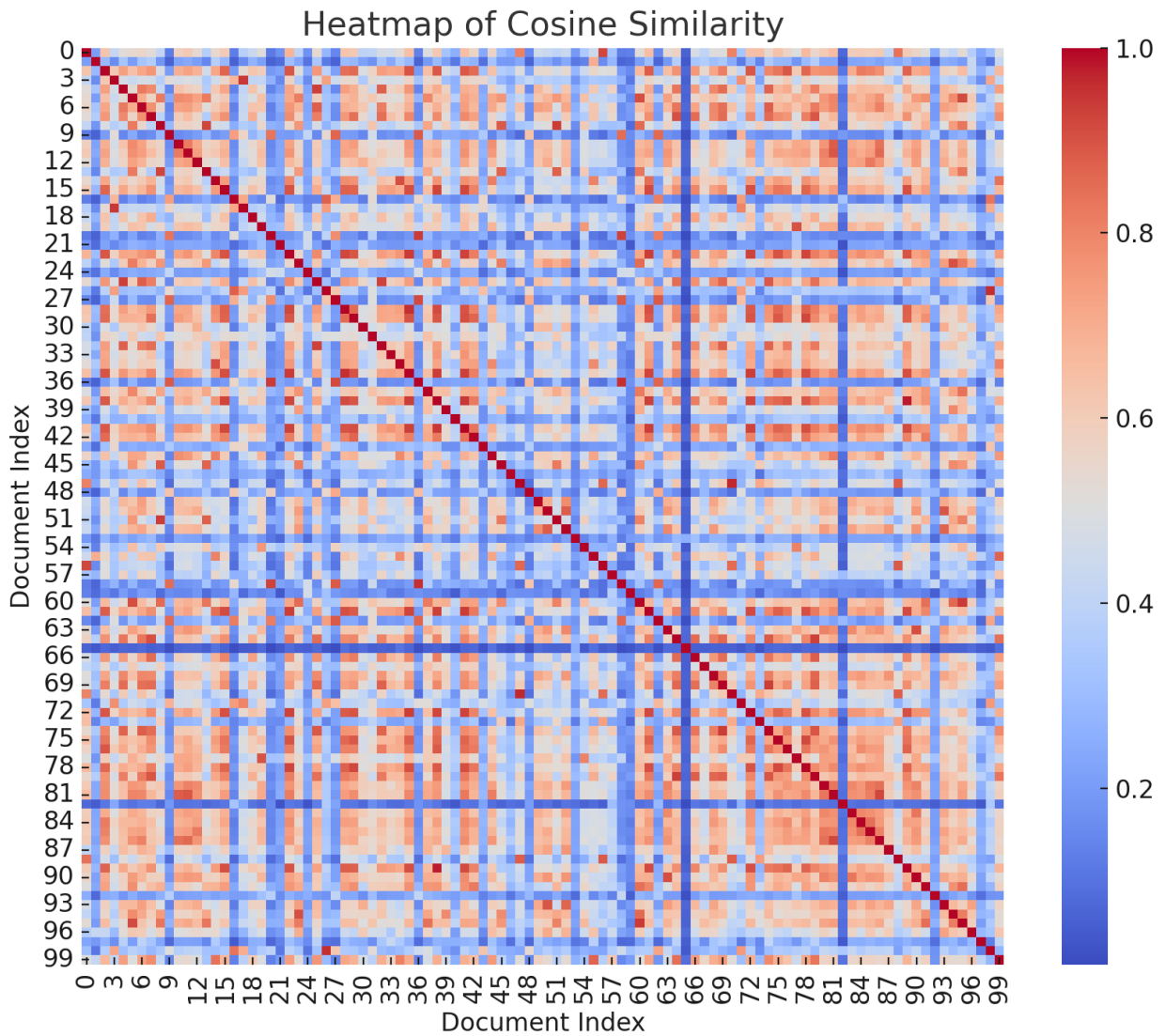
Standardhälve: 0.221

Minimaalne sarnasus: 0.010

Mediaanväärtus sarnasusest: 0.454

Maksimaalne sarnasus: 0.978





Joonis 12. Kosinussarnasuse soojuskaart 3

#### 4.4 Test number 4

Päringute arv: (200)

Tokenite arv: (~240000)

#### 4.4.1 Toortekst

Kliinilised andmed: Ülakõhuvalu. Sapikivid? Sapijuha laienemine? Kõhu ultraheliuuring. LEIUD: Pankreas on normaalse suurusega ja vanusele vastava struktuuriga; massikoldeid ei tuvastatud. Maks on normaalse suurusega ja homogeense struktuuriga; koldelisi kahjustusi ei tuvastatud. Sapipõie seinad on õhukesed ja valendik on selge. Sapijuhad ei ole laienenud. Põrn ei ole suurenenud, koldelisi kahjustusi ei esine. Neerud on normaalse suuruse, asendi ja struktuuriga; hüdroefroosi ei esine. Kusepõis on minimaalselt täidetud, anechoic sisuga. Eesnääre on ilma eripäradeta. Kõhuaort ja niudearterid on normaalse kaliibriga ja normaalse verevooluga. Retroperitoneaalselt suurenenud lümfisõlmi ei nähtu. Vaba vedelikku kõhuõõnes ei leidu.

#### 4.4.2 Tulemus

Total RDF Triples 17

<Patsient> <onSeisund> "Ülakõhuvalu"

<Patsient> <onSeisund> "Sapikivid"

<Patsient> <onSeisund> "Sapijuha laienemine"

Part 2:

<Patsient> <onTest> "Kõhu ultraheliuuring"

Part 3:

<Patsient> <onTestiTulemus> "Pankreas" "normaalne suurus"

<Patsient> <onTestiTulemus> "Pankreas" "vanusele vastav struktuur"

<Patsient> <onTestiTulemus> "Maks" "normaalne suurus"

<Patsient> <onTestiTulemus> "Maks" "homogeenne struktuur"

<Patsient> <onTestiTulemus> "Sapipõis" "õhukesed seinad"

<Patsient> <onTestiTulemus> "Sapipõis" "selge valendik"

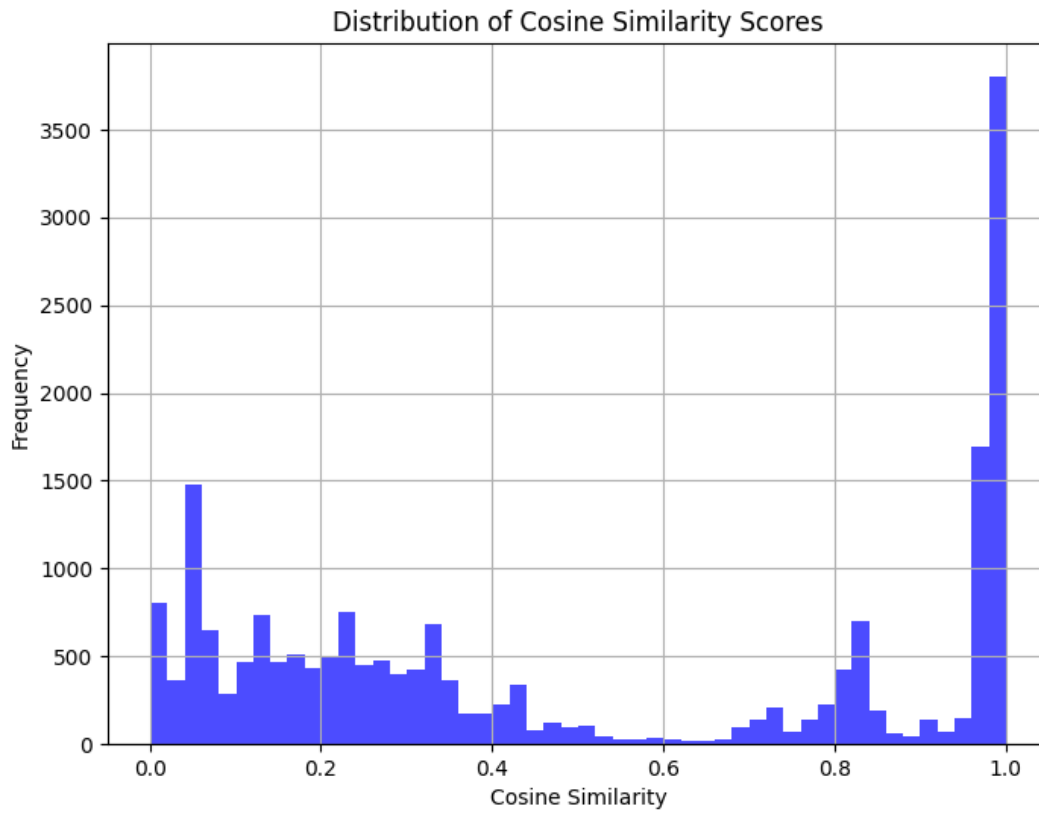
<Patsient> <onTestiTulemus> "Sapijuhad" "mitte laienenud"

<Patsient> <onTestiTulemus> "Põrn" "mitte suurenenud"

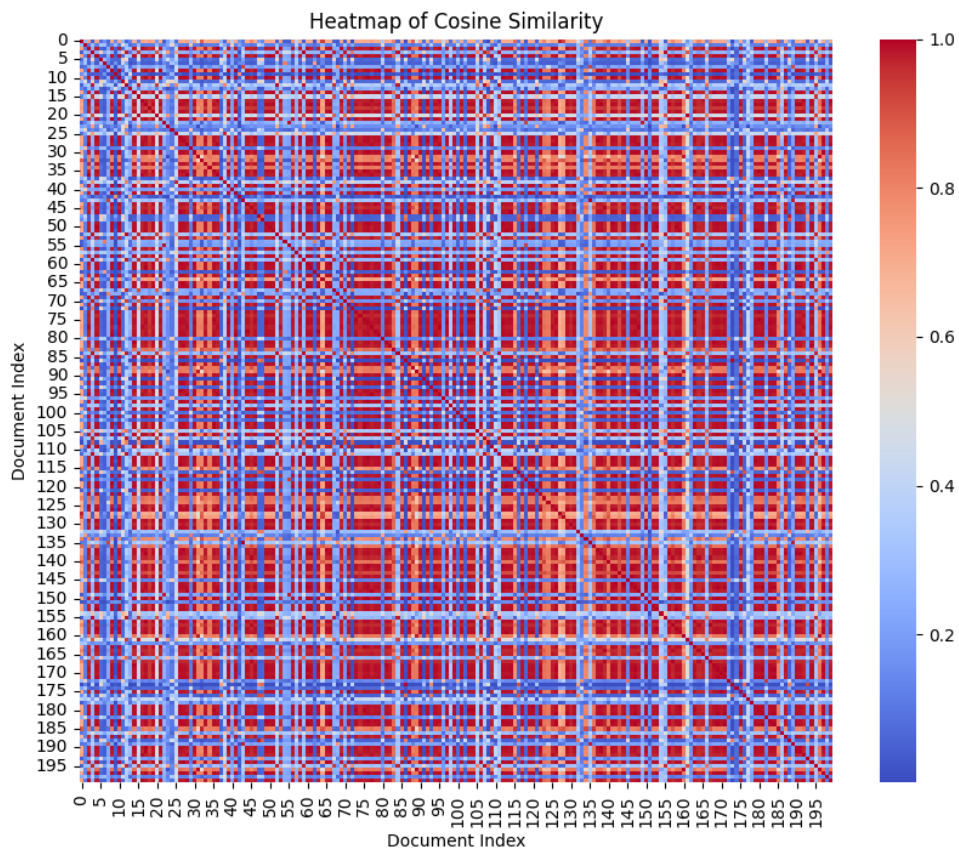
<Patsient> <onTestiTulemus> "Neerud" "normaalne suurus"

<Patsient> <onTestiTulemus> "Neerud" "normaalne asend"  
<Patsient> <onTestiTulemus> "Neerud" "normaalne struktuur"  
<Patsient> <onTestiTulemus> "Kusepõis" "minimaalselt täidetud"  
<Patsient> <onTestiTulemus> "Kusepõis" "kajavaba sisu"  
<Patsient> <onTestiTulemus> "Eesnääre" "ilma eripäradeta"  
<Patsient> <onTestiTulemus> "Kõhuaort" "normaalne kaliiber"  
<Patsient> <onTestiTulemus> "Niudearterid" "normaalne kaliiber"  
<Patsient> <onTestiTulemus> "Niudearterid" "normaalne verevool"  
<Patsient> <onTestiTulemus> "Lümfisõlmed" "mitte suurenenud"  
<Patsient> <onTestiTulemus> "Kõhuõõs" "vaba vedelik puudub"

### 4.3.3 Analüüs



Joonis 13. Koosinus Sarnastus 4



Joonis 14. Kosinussarnasuse soojuskaart 4

Keskmine sarnasuse väärtus: 0.501

Standardhälve: 0.379

Minimaalne sarnasuse väärtus: 0.011

Mediaan sarnasuse väärtus: 0.346

Maksimaalne sarnasuse väärtus: 1.000

## 4.5 Järeldused

Päringute arvu kasvuga täheldati vastuste keskmise sarnasuse suurenemist. See võib näidata, et mudel töötab suure hulga päringutega paremini. Vastuste erinevused rõhutavad vajadust sisendi standardiseerimise järele, mis aitab tagada tulemuste järjepidevuse. See hõlmab teabe esitamise ja töötlemise täpset määratlemist mudelis ning selgete juhiste väljatöötamist päringute sõnastamiseks. Tulemused näitasid, et isegi sarnaste päringute puhul võivad vastused varieeruda, mis võib olla seotud mudeli sisemiste mehhanismidega, mis põhinevad keerulistel tõenäosusjaotustel ja kontekstuaalsel analüüsil. See rõhutab ettevaatlikkust selliste mudelite kasutamisel kriitilistes rakendustes, nagu meditsiin, kus vead võivad põhjustada tõsiseid tagajärgi. Kirjete selge struktureerimine, nagu kliiniliste sümptomite, uurimistulemuste ja järelduste eraldamine, lihtsustab andmete lugemist, analüüsi ja automaatset töötlemist.

## 5 Kokkuvõte

Kokkuvõttes toob see diplomitöö esile olulised võimalused ja probleemid, mis kaasnevad GPT-3 mudeli ja kaasaegsete NLP tehnoloogiate kasutamisega meditsiinivaldkonna loomuliku keele töötlemisel. OpenAI GPT-3 mudel, eriti selle laiendus GPT-3.5 versioonile, näitab olulist võimekust meditsiinilise dokumentatsiooni automatiseerimisel ja kliiniliste otsuste toetamisel, mis võib põhjalikult muuta meditsiinipraktikat. RDF kasutamine meditsiiniliste andmete struktureerimisel parandab andmete ühilduvust ja standardiseerimist, mis on oluline eri meditsiinisüsteemide vahelise infovahetuse jaoks.

Mitmete katsete ja tulemuste analüüs tõi esile mitu suuremat probleemi. NLP töötamisel on vastuste stabiilsuse ja sarnasuse tagamine keeruline. Loomuliku keele töötlemise (NLP) valdkonnas on stabiilsuse ja sarnasuse küsimused keerukad mitmel põhjusel. Ambiguutsus ja tõlgendamisvõimalused. Meditsiinilised kirjed [5] sisaldavad sageli ebamääraseid või mittetäielikke andmeid. GPT-3 peab oskama selliseid andmeid tõlgendada, tehes eeldusi või päringuid teabe täpsustamiseks, mis võib korduvate päringute korral põhjustada vastuste erinevusi. Meditsiinilised andmed võivad sisaldada

sümptomeid, diagnoose, ravi ja muid aspekte, millest igapähele on oma unikaalsed omadused RDF-is kodeerimiseks. Sama meditsiiniline fakt võib olla kirjeldatud erinevalt, mis raskendab väljundi standardiseerimist.

RDF esituste loomise protsess GPT-3 abil ei ole andmete genereerimisel kõrge täpsuse ja stabiilsusega, mida kinnitab vastuste vahelise kosinussarnasuse analüüs. Vaatamata olemasolevatele raskustele ja potentsiaalsetele ebatäpsustele jäävad OpenAI mudelid, sealhulgas GPT-3, oluliseks tööriistaks meditsiinivaldkonnas ja neid saab kasutada mitmel otstarbel. Mudeli kasutamine nõuab täpset seadistamist, selle jõudluse pidevat hindamist ja kvaliteedi juhtimise suhtes tähelepanelikku lähenemist.

## Kasutatud allikad

[1] Language Models are Few-Shot Learners. Available: <https://arxiv.org/abs/2005.14165>

[2] BioBERT: a pre-trained biomedical language representation model for biomedical text. Available: <https://academic.oup.com/bioinformatics/article/36/4/1234/5566506>

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Available: <https://arxiv.org/abs/1603.02754>

[4] RoBERTa: A Robustly Optimized BERT Pretraining Approach. Available: <https://arxiv.org/abs/1907.11692>

[5] Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics Available: <https://ieeexplore.ieee.org/document/8322514>

[6] Ethical Dimensions of Using Artificial Intelligence in Health Care. AMA Journal of Ethics.

Available: <https://journalofethics.ama-assn.org/article/ethical-dimensions-using-artificial-intelligence-health-care/2019-02>

- [7] Amodei, D., & Hernandez, D. (2018). AI and Compute. OpenAI Blog. Available: <https://openai.com/index/ai-and-compute/>
- [8] MIMIC-III, a freely accessible critical care database. Scientific data. Available: <https://www.nature.com/articles/sdata201635>
- [9] OpenAI GPT-3: Understanding the Architecture. Available: <https://www.theaidream.com/post/openai-gpt-3-understanding-the-architecture>
- [10] Kontseptuaalne arhitektuur GPT mudelil. Available: <https://www.ruder.io/research-highlights-2020/>
- [11] The GPT-3 Architecture, on a Napkin  
Available: [https://dugas.ch/artificial\\_curiosity/GPT\\_architecture.html#](https://dugas.ch/artificial_curiosity/GPT_architecture.html#)

## Lisad

### Lisa 1 - “Automatiseeritud testimiskript”

```
def read_query_template(file_path):
    with open(file_path, 'r') as file:
        return file.read()

def send_request_to_flask(endpoint, text):
    response = requests.post(endpoint, json={'text': text})
    if response.ok:
        return response.json()
    else:
        return None

def perform_requests_and_save_to_excel(endpoint, text_template,
num_requests=200):
    results = []

    for _ in range(num_requests):
        result = send_request_to_flask(endpoint,
text_template.format(medical_text="PROMT"))
        if result is not None:
            results.append(result)

    df = pd.DataFrame(results)
    df.to_excel('file', index=False)

if __name__ == "__main__":
    FLASK_ENDPOINT = '/generate_rdf'
```



```
text_template = read_query_template()
perform_requests_and_save_to_excel(FLASK_ENDPOINT, text_template)
```

## **Lisa 2 - “Andmete puhastamise skript”**

```
wb = openpyxl.load_workbook(source_file_path)
ws = wb.active

new_wb = openpyxl.Workbook()
new_ws = new_wb.active

def clean_text(text):
    return re.sub(r'^\w\s', '', text)

for idx, row in enumerate(ws.iter_rows(min_row=2, min_col=1,
max_col=1), start=2):
    for cell in row:
        cleaned_text = clean_text(cell.value) if cell.value else ''
        new_ws[f'A{idx}'].value = cleaned_text

if new_ws.max_row >= 1:
    new_ws.delete_rows(1)

new_wb.save(new_file_path)
```

## **Lisa 3 - “TF-IDF ja kosinussarnasuse analüüsi skript tekstiliste andmete jaoks”**

```
def plot_cosine_similarity_histogram(file_path):
    data = pd.read_excel(file_path)

    upper_triangle_values = data.where(np.triu(np.ones(data.shape),
k=1).astype(bool)).stack()

    pyplot.figure(figsize=(8, 6))
    pyplot.hist(upper_triangle_values, bins=50, color='blue',
alpha=0.7)
```

```
pyplot.title('Distribution of Cosine Similarity Scores')
pyplot.xlabel('Cosine Similarity')
pyplot.ylabel('Frequency')
pyplot.grid(True)
pyplot.show()
```

```
plot_cosine_similarity_histogram(file_path)
```

## **Lisa 4 - “Tekstidokumentide kosinussarnasuse soojuskaardi visualiseerimise skript”**

```
def plot_cosine_similarity_heatmap(file_path):
    data = pd.read_excel(file_path)

    pyplot.figure(figsize=(10, 8))
    heatmap = sns.heatmap(data, cmap='coolwarm', annot=False)
    pyplot.title('Heatmap of Cosine Similarity')
    pyplot.xlabel('Document Index')
    pyplot.ylabel('Document Index')
    pyplot.show()

plot_cosine_similarity_heatmap(file_path)
```

## **Lisa 5 - “GPT-3.5 Turbo päringu seadistused.”**

```
response = openai.completions.create(
    model="gpt-3.5-turbo-instruct",
    prompt=prompt,
    temperature=0.6,
    max_tokens=3000,
    top_p=1.0,
    frequency_penalty=0.0,
    presence_penalty=0.0
)
```