

МОНГОЛ УЛСЫН ИХ СУРГУУЛЬ  
МЭДЭЭЛЛИЙН ТЕХНОЛОГИ, ЭЛЕКТРОНИКИЙН СУРГУУЛЬ  
МЭДЭЭЛЭЛ, КОМПЬЮТЕРЫН УХААНЫ ТЭНХИМ

Бадарчийн Бат-Энх

Үйлдвэрлэлийн дадлага  
(Internship report)

Компьютерын Ухаан (D061301)  
Бакалаврын судалгааны ажил

Улаанбаатар

2026 оны 2 дугаар сар

МОНГОЛ УЛСЫН ИХ СУРГУУЛЬ  
МЭДЭЭЛЛИЙН ТЕХНОЛОГИ, ЭЛЕКТРОНИКИЙН СУРГУУЛЬ  
МЭДЭЭЛЭЛ, КОМПЬЮТЕРЫН УХААНЫ ТЭНХИМ

Үйлдвэрлэлийн дадлага  
(Internship report)

Компьютерын Ухаан (D061301)  
Бакалаврын судалгааны ажил

Удирдагч:	_____	У. Нямбаяр
Хамтран удирдагч:	_____	
Гүйцэтгэсэн:	_____	Б.Бат-Энх (22B1NUM7226)

Улаанбаатар  
2026 оны 2 дугаар сар

# Зохиогчийн баталгаа

Миний бие Бадарчийн Бат-Энх "Үйлдвэрлэлийн дадлага" сэдэвтэй судалгааны ажлыг гүйцэтгэсэн болохыг зарлаж дараах зүйлсийг баталж байна:

- Ажил нь бүхэлдээ Монгол Улсын Их Сургуульд дээд боловсролын зэрэг горилохоор дэвшүүлсэн болно.
- Энэ ажлын аль нэг хэсгийг эсвэл бүхлээр нь ямар нэг их, дээд сургуулийн зэрэг горилохоор оруулж байгаагүй болно.
- Бусдын хийсэн ажлаас хуулбарлаагүй, эшлэл, зүүлтийг зохистой хийсэн болно.
- Ажлыг зохиогч би хийсэн ба миний хийсэн ажил, бусдын дэмжлэгийг дипломын ажилд тодорхой тусгасан болно.
- Ажилд тусалсан бүх эх сурвалжид талархаж байна.

Гарын үсэг: \_\_\_\_\_

Огноо: \_\_\_\_\_

## ГАРЧИГ

УДИРТГАЛ .....	1
1. БАЙГУУЛЛАГЫН ТАНИЛЦУУЛГА .....	2
1.1 Unitel Group .....	2
1.2 Unitel Group-ийн үйлчилгээ .....	2
1.3 Дадлагын албан тушаал .....	3
2. СУДАЛГАА .....	4
2.1 Agentic AI .....	4
2.2 Model Context Protocol .....	5
2.3 Huawei Cloud Stack .....	5
2.4 Langfuse .....	6
2.5 Gemini .....	7
2.6 AWS Cohere .....	7
3. ХЭРЭГЖҮҮЛЭЛТ .....	8
3.1 Enterprise Resource Planning MCP .....	8
3.2 Information Retrieval MCP .....	8
3.3 Agent Skill .....	11
3.4 TV content data entry automation .....	11
3.5 Huawei Cloud Stack .....	11
3.6 Deployment infrastructure .....	11
4. ҮР ДҮН .....	12
ДҮГНЭЛТ .....	13
ХАВСРАЛТ .....	15
А. ....	16
В. ....	17

## ЗУРГИЙН ЖАГСААЛТ

2.1	Workflow .....	4
2.2	Agentic .....	4
2.3	HCS дээрх системийн зохион байгуулалт .....	5
2.4	VPN холболтын диаграм .....	6

## ХҮСНЭГТИЙН ЖАГСААЛТ

## Кодын жагсаалт

3.1	Qdrant client initialization . . . . .	9
3.2	Qdrant client вектор хайлт . . . . .	10

## УДИРТГАЛ

Баримт нь МКУТ-ын анги дипломын ажил бичихэд туслах загвар бөгөөд энд ашигласан командуудыг өөрийн баримтад тохируулан ашиглана.

Жич: Дипломын ажлыг бичихдээ удиртгалд сэдвийг гүйцэтгэх болсон үндэслэл, зорилго, зорилтыг заавал оруулан бичнэ.

LaTeX-ийг сурахдаа энэ загварын командуудыг жишээ болгон ашиглаж болохоос гадна [https://www.sharelatex.com/learn/Learn\\_LaTeX\\_in\\_30\\_minutes](https://www.sharelatex.com/learn/Learn_LaTeX_in_30_minutes) холбоосоос 30 минутад сурч болно.

LaTeX компайлрыг өөрийн компьютерт суулгах бол MikTeX-ийг интернетээс татаж суулгана. Энэ нь зөвхөн компайлар ба засварлагч давхар суулгах шаардлагатай. Хамгийн нийтлэг хэрэглэгддэг засварлагч нь Texmaker юм.

Хэрэв өөрийн компьютерт суулгаж болохгүй бол <https://www.overleaf.com>-д бүртгүүлээд онлайнаар ажиллаж болно.



# 1. БАЙГУУЛЛАГЫН ТАНИЛЦУУЛГА

## 1.1 Unitel Group

Юнител групп нь Монгол Улсад үйл ажиллагаа явуулдаг харилцаа холбоо, мэдээллийн технологийн компани юм. Тус компани нь 2005 онд үүсгэн байгуулагдаж, 2006 оноос GSM үүрэн холбооны үйлчилгээ үзүүлж эхэлсэн. Үүсгэн байгуулагдсанаасаа хойш Юнител групп нь Монгол Улсын харилцаа холбооны салбарын хөгжилд чухал үүрэг гүйцэтгэж, дэвшилтэт технологийг тасралтгүй нэвтрүүлж ирсэн.

Юнител нь Монгол Улсад 4G LTE болон 5G технологийг нэвтрүүлсэн анхдагч компаниудын нэг бөгөөд хот, хөдөөгийн харилцаа холбооны дэд бүтцийг өргөжүүлэх, өндөр хурдны интернэтийн хүртээмжийг нэмэгдүүлэхээр ажиллаж байна. Тус компани нь MCS Holding н бүрэлдэхүүнд багтдаг.

## 1.2 Unitel Group-ийн үйлчилгээ

Юнител групп нь хэрэглэгчдэд дараах үндсэн үйлчилгээнүүдийг үзүүлдэг. Үүнд:

- Үүрэн холбооны үйлчилгээ (дуу, мессеж, дата)
- Өргөн зурвасын интернэт үйлчилгээ
- Суурин утасны үйлчилгээ
- IPTV болон дижитал телевизийн үйлчилгээ
- Стриминг болон контент үйлчилгээ
- Хиймэл дагуулын холбооны үйлчилгээ

Эдгээр үйлчилгээнүүд нь хувь хэрэглэгч болон байгууллагын хэрэгцээнд нийцсэн байдлаар санал болгогддог бөгөөд Монгол Улсын мэдээллийн технологи, дижитал шилжилтийг дэмжихэд чиглэсэн байдаг.

### **1.3 Дадлагын албан тушаал**

Миний бие энэ дадлагын хүрээнд Unitel компанийн Дижитал трансформейшний газарт Дата сайн­тис­тын ажлын байранд ажилласан. Манай хэлтсийн хувьд компани дахь бусад баг, хэлтэстэй хамтран AI системүүдийн нэвтрүүлэлт, автоматжуулалтын ажлуудад түлхүү чиглэсэн ажил гүйцэтгэдэг.

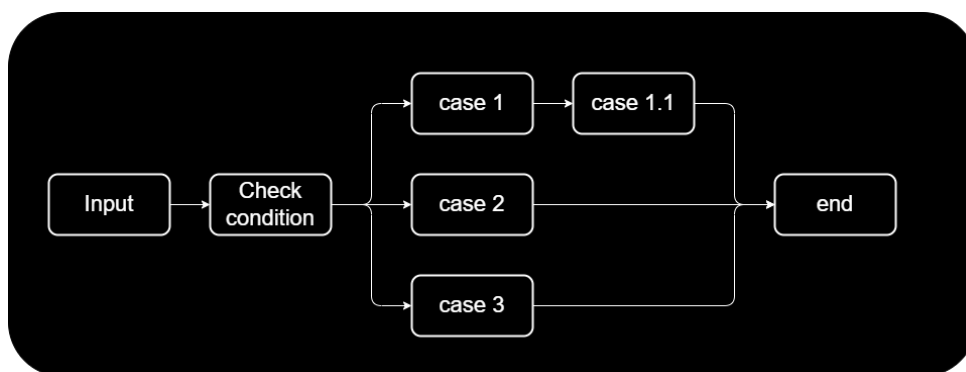
Дадлагын ажлын хугацаанд миний бие ажилтанд зориулсан AI туслах, Call center-ийн ажилтны AI туслах, TV content автомат дата оруулагч болон хэлтсийн Development ба Pro­duction орчны зохион байгуулалтуудыг хариуцан ажилласан.

## 2. СУДАЛГАА

Unitel Group нь дэлхийн хиймэл оюуны ухааны хувьсгалтай ижил түвшинд хөгжих зорилготой. Иймд шинэ технологи болон архитектуруудыг компаны хэрэгцээнд зориулан байнга нэвтрүүлэн ашигладаг.

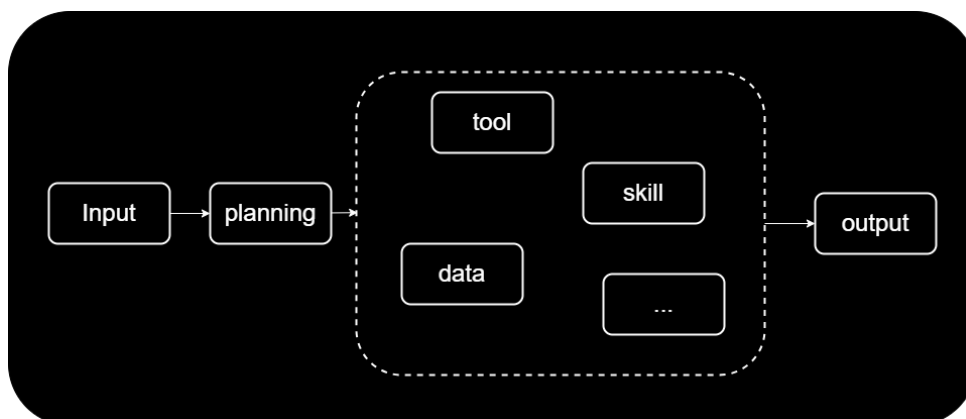
### 2.1 Agentic AI

Agentic architecture нь нэг ёсондоо өөрийгөө тодорхойлдог workflow юм [6] [4]. Large Language Model (LLM)-ийн огцом хөгжил нь энэ архитектурын динамик байдлыг хангаж комплекс ажлуудыг програмчлалын аргаар шийдэх боломжийг олгосон.



Зураг 2.1: Workflow

Workflow нь статик програмчлагдсан бөгөөд зорилго болон зорилго биелүүлэхэд хэрэгцээтэй зүйлсийг тодорхойлох механизм байхгүй.



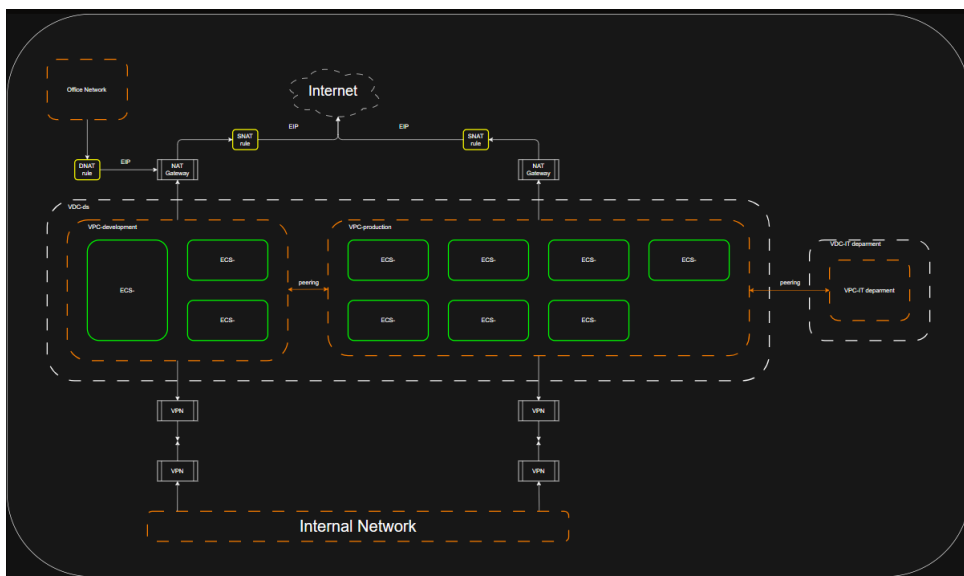
Зураг 2.2: Agentic

## 2.2 Model Context Protocol

Antropic компанийн хөгжүүлсэн тус протокол нь LLM-д tool call хийх боломж олгох хамгийн уян хатан технологи юм. Model Context Protocol ашигласнаар LLM нь гадны өгөгдөлтэй харицах стандарт харилцааны протоколтой болсон. Үүний үр дүнд LLM-ийн модель болгон өөрийн гэсэн tool-тэй байхын оронд нэг tool-ийг олон модель зэрэг ашиглах, өөр хүмүүсийн бэлтгэсэн tool-ийг ашиглах зэрэг боломжууд үүссэн [8].

## 2.3 Huawei Cloud Stack

Huawei Cloud Stack (HCS) нь Data Center удирдлагын систем бөгөөд Unitel компани өөрийн Data Center-ээ тус системээр удирддаг бөгөөд нэвтрүүлэлтийн болон хөгжүүлэлт ажлуудыг удирдахад тус системийн тухай мэдлэгтэй байх шаардлагатай байдаг [5].



Зураг 2.3: HCS дээрх системийн зохион байгуулалт

### 2.3.1 Elastic Compute Service

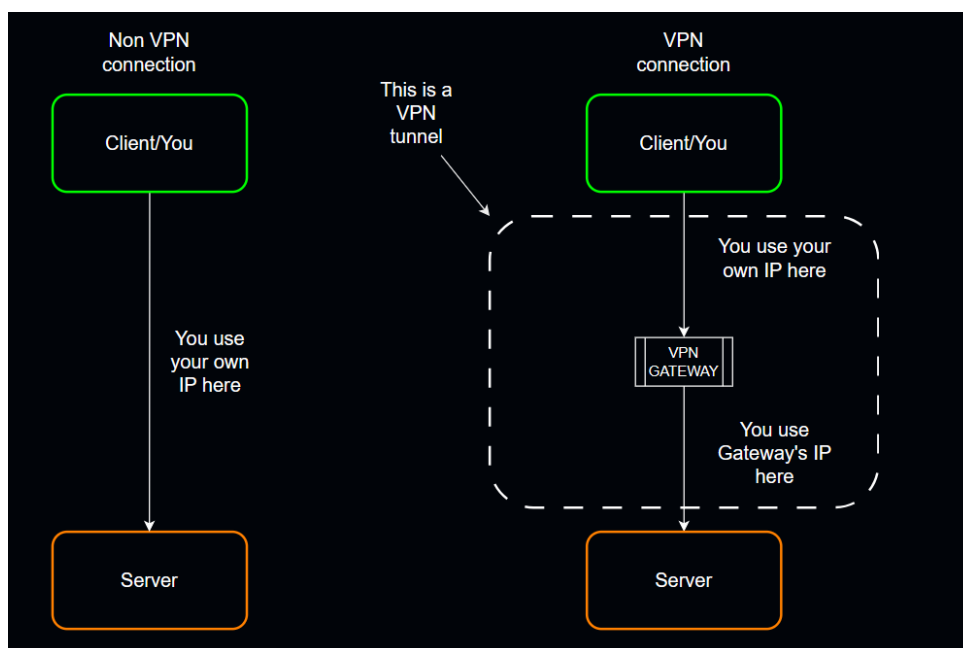
HCS-ийн Elastic Compute Service (ECS) нь AWS-ийн EC2-тэй дүйх бөгөөд аливаа сүлжээний subnet дээр орших виртуал машин юм [10][5].

### 2.3.2 Virtual Private Cloud

HCS дотор орших тусгаарлагдсан Data Center-ийн дотоод сүлжээ юм. VPC-г зөв зохион байгуулах нь цаашид хандалтыг хязгаарлах, бусад гадны сүлжээтэй харилцах гэх мэт боломжуудыг олгоно [5].

### 2.3.3 Virtual Private Network

VPC-ийн дотоод сүлжээ нь гадны дотоод сүлжээтэй холбогдохын тулд VPN холболт үүсгэх шаардлагатай байдаг [5].



Зураг 2.4: VPN холболтын диаграм

## 2.4 Langfuse

Langfuse нь LLM дээр суурилсан системүүдийн ажиглалт, хяналт болон гүйцэтгэлийн шинжилгээнд зориулагдсан open-source платформ юм [7]. Agentic AI архитектурын хувьд олон шатлалт reasoning, tool call, decision making процессууд явагддаг тул эдгээрийг бүртгэн хадгалах шаардлага үүсдэг. Langfuse нь prompt, response, latency, token usage зэрэг мэдээллүүдийг цуглуулж, LLM-ийн үйл ажиллагааг бодит орчинд хянах боломжийг олгодог.

Unitel Group-ийн хувьд LLM-д суурилсан системүүдийг үйлдвэрлэлийн орчинд ашиглах

үед ил тод байдал, алдааг илрүүлэх, гүйцэтгэлийг оновчлох нь маш чухал. Langfuse нь энэ шаардлагыг хангах хяналтын давхаргыг бүрдүүлж өгдөг.

## 2.5 Gemini

Gemini нь Google компанийн хөгжүүлсэн Large Language Model юм [2]. Тус модельд embeddings үүсгэх функциональ 2025 онд шинээр нэмэгдсэн.

Gemini embedding нь текстийг 3072 хэмжээтэй вектор болгон хувиргадаг. Agentic AI архитектурын хүрээнд embeddings-ийг ашигласнаар систем нь хадгалагдсан мэдлэгтэй уялдуулан шийдвэр гаргах боломж бүрддэг. Энэ нь workflow-оос agentic архитектурт шилжихэд шаардлагатай контекстийн ойлголтыг хангана.

## 2.6 AWS Cohere

Cohere нь Large Language Model болон embeddings үйлчилгээ үзүүлэгч бөгөөд AWS орчинд ашиглах боломжтойгоороо онцлог юм [1].

Unitel Group саяхныг хүртэл Cohere Multilingual v3 embedding моделийг ашигладаг байсан бөгөөд 2026 онд хийсэн туршилт дээр үндэслэн Cohere-ийг fallback модель, харин gemini-г үндсэн модель болгон шилжүүлсэн.

## 3. ХЭРЭГЖҮҮЛЭЛТ

### 3.1 Enterprise Resource Planning MCP

### 3.2 Information Retrieval MCP

Дадлагын хугацаанд Information Retrieval MCP tool-ийг хариуцсан ажилтны хувьд дараах ажлуудыг хийсэн.

#### 3.2.1 Вектор өгөгдлийн сан солих

Qdrant нь Open-Source вектор өгөгдлийн сан юм [9]. Unitel Chatbot болон бусад RAG технологи ашигладаг системүүдэд AstraDB-ийн Cloud вектор өгөгдлийн санг ашигладаг байсан бол энэ жилээс эхлэн Self-Hosted хувилбар луу шилжих ажил хийгдэж хэд хэдэн вектор өгөгдлийн санг харьцуулснаас Qdrant хамгийн тохиромжтой гэх дүгнэлтэд хүрсэн. Үүнд харгалзаж үзсэн үзүүлэлтүүдийг дурдвал:

- Retrieval latency
- Self-Host үнийн санал
- Өөрийн гэсэн User Interface
- Вектор өгөгдлийн сан дээрх хийгдэж болох үйлдлүүд:
  - Metadata filtering
  - Tenant based separation
  - Access control

Иймд Information Retrieval MCP tool-ийн ашигладаг вектор өгөгдлийн санг солих ажлыг дараах байдлаар солисон.

```
1
2 class Qdrant:
3     def __init__(self):
4         settings = get_settings()
5         self.gemini_embedder = GoogleEmbedder().google_embeddings
6         self.cohere_embedder = AWSEmbedder().aws_embeddings
7         self.client = QdrantClient(
8             url=settings.QDRANT_CLIENT_URL,
9             api_key=settings.QDRANT_API_KEY,
10            timeout=20.0 # 60 second timeout for Qdrant
11                operations
12        )
13        self.QDRANT_PROD_MODE=settings.QDRANT_PROD_MODE
```

Код 3.1: Qdrant client initialization



```
1      async def vector_search(self, query: str, filter=None, tenant_id=
2          None, embedding_name=None, top_k: int=10):
3
4          collection_name = self._qdrant_collection_name(tenant_id=
5              tenant_id, embedding_name=embedding_name)
6
7          #embded query
8
9          if embedding_name == "gemini":
10
11              vec = await self.gemini_embedder.aembed_query(query)
12
13          else:
14
15              vec = await self.cohere_embedder.aembed_query(query)
16
17          hits = self.client.query_points(
18              collection_name=collection_name,
19              query=vec,
20              query_filter=filter,
21              limit=top_k,
22          )
23
24          return hits
25
26      def __call__(self, query, filter, top_k=10):
27
28          self.vector_search(query=query, filter=filter, top_k=top_k)
```

Код 3.2: Qdrant client вектор хайлт

### 3.2.2 AWS Cohere

Энэ онд хийгдэж эхэлсэн ажлуудад вектор өгөгдлийн санг шинэчлэхээс гадна embed хийж буй моделиудыг шинэчлэх мөн fallback модель сонгох ажил багтсан. Өмнө нь AWS-ийн Cohere Multilingual v3 [1] модель Монгол хэл дээр хамгийн сайн үр дүн үзүүлсэн тул сүүлийн 2 жилийн турш сонгон ашигласан. Харин хамгийн сүүлд хийсэн туршилтын үр дүнд Gemini Embedding-ийг main моделиор харин Cohere Multilingual v3 fallback моделиор сонгогдсон.

### 3.2.3 Gemini Embedding

Google-ийн хөгжүүлсэн gemini-embedding-001 модель нь олон нийтэд 2025 оноос эхлэн нээлттэй болсон [3]. Харин Unitel компаны дотооддоо хийсэн туршилтын үр дүнд Монгол хэл дээр хамгийн сайн үр дүн үзүүлсэн. Туршилтаар embedding моделиудыг:

- Хайлтын хурд
- Хайлтын оновчтой байдал

гэсэн хоёр үзүүлэлтүүдэд тулгуурлан дүгнэсэн.

## 3.3 Agent Skill

## 3.4 TV content data entry automation

## 3.5 Huawei Cloud Stack

### 3.5.1 Development server

### 3.5.2 Production server

### 3.5.3 Peering

### 3.5.4 VPN connection

## 3.6 Deployment infrastructure

## **4. ҮР ДҮН**

## Дүгнэлт

nice

# Bibliography

- [1] Cohere. Cohere documentation. <https://docs.cohere.com>, 2026. Accessed: 2026-02-03.
- [2] Google. Gemini api documentation. <https://ai.google.dev/gemini-api/docs>, 2026. Accessed: 2026-02-06.
- [3] Google AI. Gemini api documentation: Embeddings. <https://ai.google.dev/gemini-api/docs/embeddings>, 2026. Accessed: 2026-02-03.
- [4] Google Cloud. What is agentic ai. <https://cloud.google.com/discover/what-is-agentic-ai>, 2026. Accessed: 2026-02-03.
- [5] Huawei. Huawei cloud stack documentation. <https://support.huawei.com/enterprise/en/cloud-computing/huawei-cloud-stack-pid-23864287>, 2026. Accessed: 2026-02-03.
- [6] IBM. What is agentic ai? <https://www.ibm.com/think/topics/agentic-ai>, 2026. Accessed: 2026-02-02.
- [7] Langfuse. Langfuse documentation. <https://langfuse.com/docs>, 2026. Accessed: 2026-02-03.
- [8] Model Context Protocol. Getting started with model context protocol. <https://modelcontextprotocol.io/docs/getting-started/intro>, 2026. Accessed: 2026-01-29.

- [9] Qdrant. Qdrant documentation. <https://qdrant.tech/documentation/>, 2026. Accessed: 2026-01-29.
- [10] Amazon Web Services. Amazon web service documentation. <https://docs.aws.amazon.com>, 2026. Accessed: 2026-02-06.

**A.**

Хавсралтын агуулга

## B.

```
1 import numpy as np
2
3 def incmatrix(genl1,genl2):
4     m = len(genl1)
5     n = len(genl2)
6     M = None #to become the incidence matrix
7     VT = np.zeros((n*m,1), int) #dummy variable
8
9     #compute the bitwise xor matrix
10    M1 = bitxormatrix(genl1)
11    M2 = np.triu(bitxormatrix(genl2),1)
12
13    for i in range(m-1):
14        for j in range(i+1, m):
15            [r,c] = np.where(M2 == M1[i,j])
16            for k in range(len(r)):
17                VT[(i)*n + r[k]] = 1;
18                VT[(i)*n + c[k]] = 1;
19                VT[(j)*n + r[k]] = 1;
20                VT[(j)*n + c[k]] = 1;
21
22            if M is None:
23                M = np.copy(VT)
24            else:
25                M = np.concatenate((M, VT), 1)
26
27            VT = np.zeros((n*m,1), int)
28
29    return M
```