



OPENCLASSROOMS

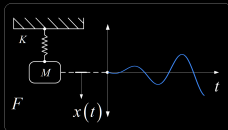
Formation **Ingénieur Machine Learning**

Projet: Définissez votre **stratégie d'apprentissage**

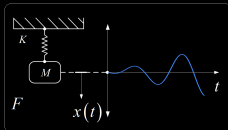
5 Janvier 2023

thomas.durandtexte@protonmail.com

Mon **parcours**



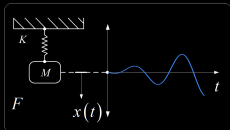
Diplôme ingénieur



Diplôme ingénieur

Master international



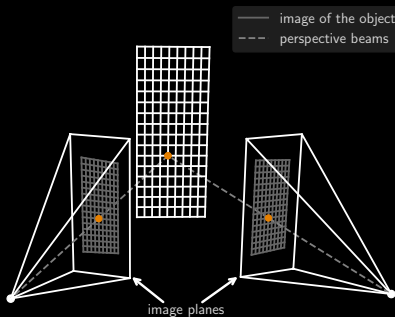


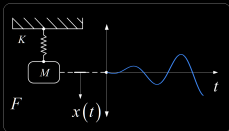
Diplôme ingénieur

Master international



Thèse

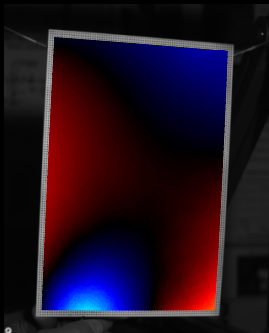


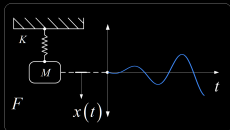


Diplôme ingénieur

Master international

Thèse



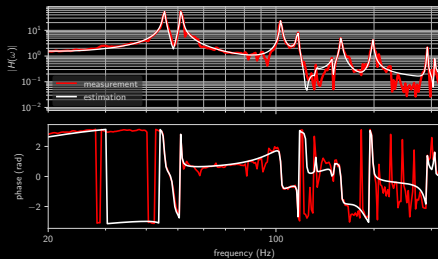


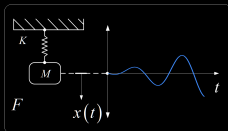
Diplôme ingénieur

Master international



Thèse



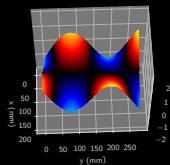
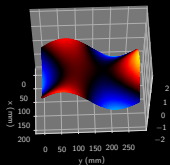
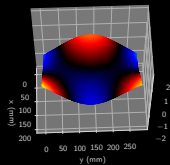
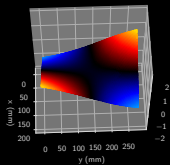


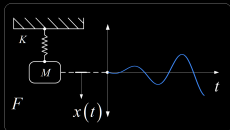
Diplôme ingénieur

Master international



Thèse

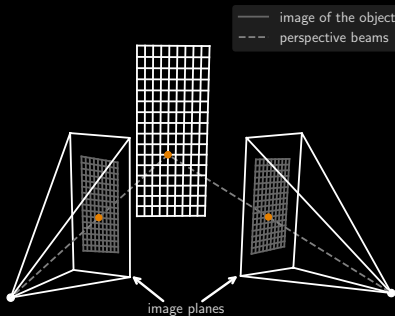


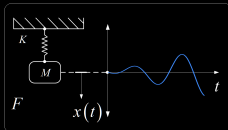


Diplôme ingénieur

Master international

Thèse





Diplôme ingénieur

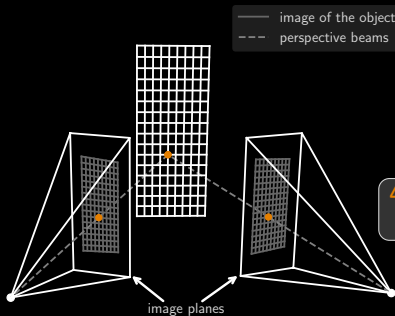
Master international

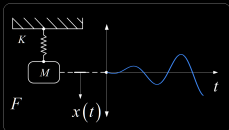


Thèse

**4 ans recherche
post-doctorale**

Imagerie
infra-rouge
Amortissement
vibratoire
Céramiques
électro-caloriques





Diplôme ingénieur

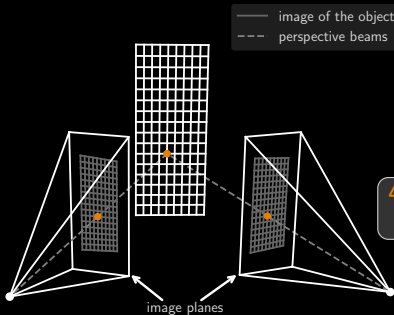
Master international



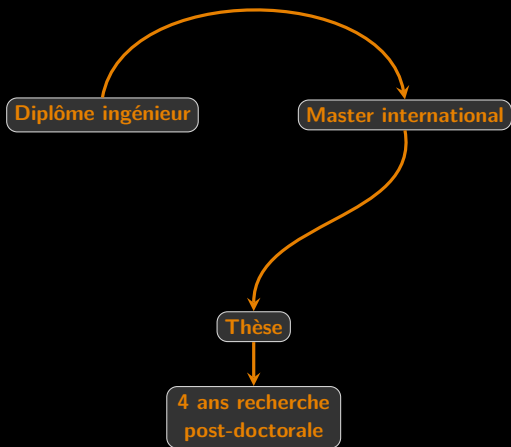
Thèse

4 ans recherche
post-doctorale

Imagerie
infra-rouge
Amortissement
vibratoire
Céramiques
électro-caloriques



MACHINE
LEARNING



Pourquoi une nouvelle formation ?

Dans **quel but** ?



Plaisir d'**apprendre**





Élargir mon champ
de **compétences**

Plaisir d'**apprendre**





Élargir mon champ
de **compétences**

Plaisir d'**apprendre**



Accéder à
de nouvelles
opportunités





Élargir mon champ
de **compétences**

Plaisir d'**apprendre**



Accéder à
de nouvelles
opportunités



CHANCE



Plaisir d'**apprendre**



Élargir mon champ
de **compétences**



Accéder à
de nouvelles
opportunités



Création
micro-entreprise
(freelance)



Pourquoi OpenClassrooms ?



**Formations
reconnues**



**Formations
reconnues**



**Mentorat
45 min/sem.**



**Formations
reconnues**



**Mentorat
45 min/sem.**



**Accès à une
communauté**



**Formations
reconnues**



**Mentorat
45 min/sem.**

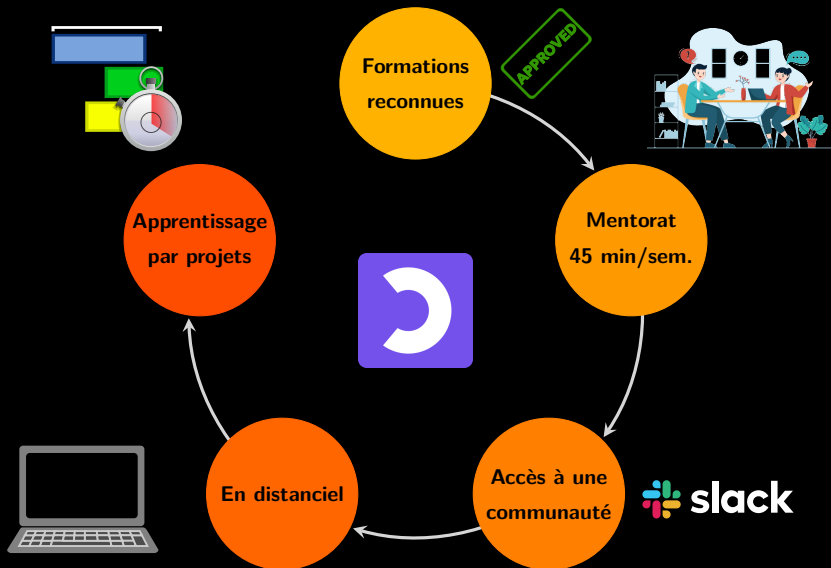


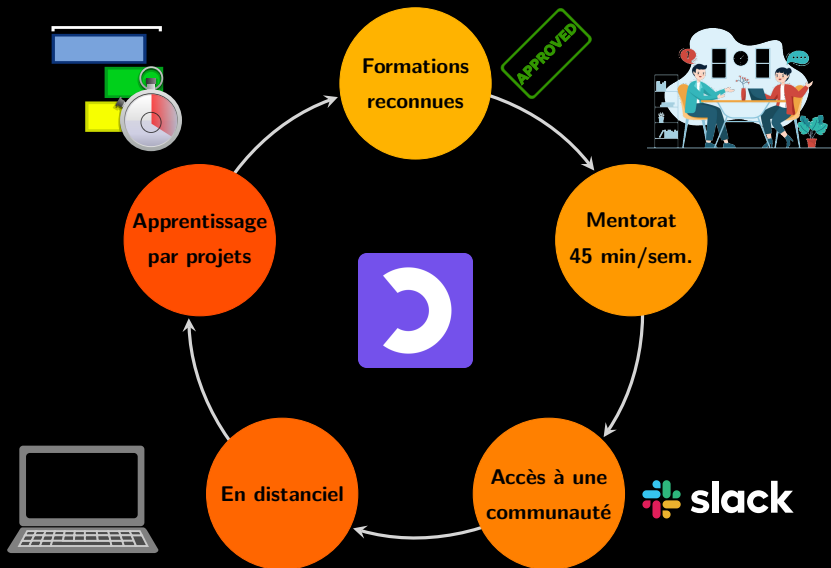
**Accès à une
communauté**



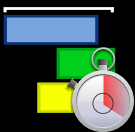
En distanciel



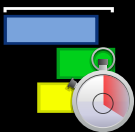




Les projets



Projets



Projets

1

2

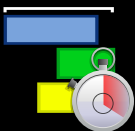
3

4

5

6

7



Projets

Concevez une application
au service de la santé publique

- **Gestion** de **données**
incomplètes
- **Analyse statistique**
uni/multivariée
- **Communication** sur les
résultats

1

2

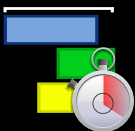
3

4

5

6

7



Projets

Santé

1

Concevez une application
au service de la santé publique

- **Gestion** de **données**
incomplètes
- **Analyse statistique**
uni/multivariée
- **Communication** sur les
résultats

2

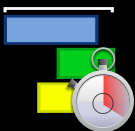
3

4

5

6

7



Projets

Santé

2

3

4

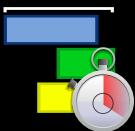
5

6

7

Anticipez les besoins
en consommation de bâtiments

- Mise en place d'un **modèle** d'apprentissage **supervisé**
- **Transformation** des **variables** pertinentes
- **Adaptation** des hyperparamètres
- **Évaluation** des **performances** du modèle



Projets

Santé

Énergie

2

3

4

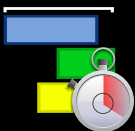
5

6

7

Anticipez les besoins
en consommation de bâtiments

- Mise en place d'un **modèle** d'apprentissage **supervisé**
- **Transformation** des **variables** pertinentes
- **Adaptation** des hyperparamètres
- **Évaluation** des **performances** du modèle



Projets

Santé

Énergie

3

4

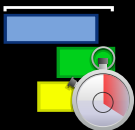
5

6

7

Segmentez des clients
d'un site e-commerce

- Mise en place d'un **modèle** d'apprentissage **non-supervisé**
- **Transformation** des **variables** pertinentes
- **Adaptation** des hyperparamètres
- **Évaluation** des **performances** du modèle



3

4

5

6

7

Projets

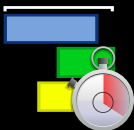
Segmentez des clients
d'un site e-commerce

- Mise en place d'un **modèle** d'apprentissage **non-supervisé**
- **Transformation** des **variables** pertinentes
- **Adaptation** des hyperparamètres
- **Évaluation** des **performances** du modèle

Santé

Énergie

Segmenter



4

5

6

7

Projets

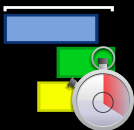
Catégorisez automatiquement
des questions

- **Gestion** de **données** à grandes dimensions
- **Prétraitement** des **données non structurées**
- Mise en œuvre des techniques de **réduction de dimension**
- Mise en œuvre des techniques d'**extraction de features** pour des données non structurées

Santé

Énergie

Segmenter



4

5

6

7

Projets

Catégorisez automatiquement
des questions

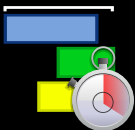
- **Gestion** de **données** à grandes dimensions
- **Prétraitement** des **données non structurées**
- Mise en œuvre des techniques de **réduction de dimension**
- Mise en œuvre des techniques d'**extraction de features** pour des données non structurées

Santé

Énergie

Segmenter

Catégoriser



Projets

Classez des images à l'aide d'algorithmes de Deep Learning

- Mise en place d'un **modèle** de **Deep Learning**
- **Sélection** d'un modèle d'apprentissage **adapté**
- **Transformation** des **variables** pertinentes
- **Évaluation** des **performances** du modèle

Santé

Énergie

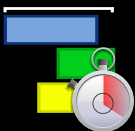
Segmenter

Catégoriser

5

6

7



Projets

Classez des images à l'aide d'algorithmes de Deep Learning

- Mise en place d'un **modèle** de **Deep Learning**
- **Sélection** d'un modèle d'apprentissage **adapté**
- **Transformation** des **variables** pertinentes
- **Évaluation** des **performances** du modèle

5

6

7

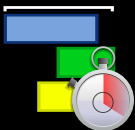
Santé

Énergie

Segmenter

Catégoriser

Classer/images



Projets

Développez une
preuve de concept

- **Réalisation** d'une **veille** sur les évolutions de la **Data Science**
- **Développement** d'une preuve de concept
- **Identification** des méthodes "état de l'art"
- **Identification** des **sources** d'informations **fiables** et **pertinentes**

Santé

Énergie

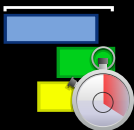
Segmenter

Catégoriser

Classer/images

6

7



Projets

Développez une
preuve de concept

- **Réalisation** d'une **veille** sur les évolutions de la **Data Science**
- **Développement** d'une preuve de concept
- **Identification** des méthodes "état de l'art"
- **Identification** des **sources** d'informations **fiables** et **pertinentes**

Santé

Énergie

Segmenter

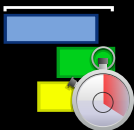
Catégoriser

Classer/images

PoC

6

7



Projets

Participez à une
compétition Kaggle !

- **Présentation** de code aux standards **PEP 8**
- **Gestion** des **versions de code**
- **Rédaction** d'une **note méthodologique**
- **Enrichissement** des **réalisations** d'autres membres de la **communauté** de professionnels

Santé

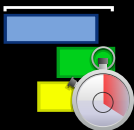
Énergie

Segmenter

Catégoriser

Classer/images

PoC



Projets

Participez à une
compétition Kaggle !

- **Présentation** de code aux standards **PEP 8**
- **Gestion** des **versions de code**
- **Rédaction** d'une **note méthodologique**
- **Enrichissement** des **réalisations** d'autres membres de la **communauté** de professionnels

Santé

Énergie

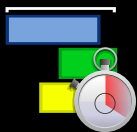
Segmenter

Catégoriser

Classer/images

PoC

Kaggle



Projets

Santé

Énergie

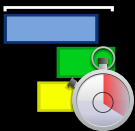
Segmenter

Catégoriser

Classer/images

PoC

Kaggle



Projets

Les plus **motivants**

Santé

Énergie

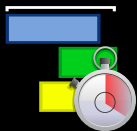
Segmenter

Catégoriser

Classer/images

PoC

Kaggle



Projets

Les plus **motivants**

Santé

Énergie

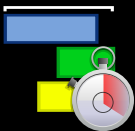
Classifier/images

Segmenter

Catégoriser

PoC

Kaggle



Projets

Les plus **motivants**

Santé

Énergie

Classifier/images

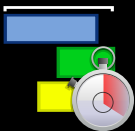
Difficiles (à priori)

Segmenter

Catégoriser

PoC

Kaggle



Projets

Les plus **motivants**

Santé

Énergie

Classifier/images

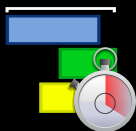
Difficiles (à priori)

Segmenter

Catégoriser

Kaggle

PoC



Projets

Les plus **motivants**

Santé

Énergie

Classifier/images

Difficiles (à priori)

Segmenter

Catégoriser

Kaggle

?

PoC

Planification

Notebook pour estimer les **dates prévisionnelles** de soutenances :

Import des packages et chargement des données

Les données correspondent à un temps estimatif de travail des différents projets

```
import pandas as pd
import numpy as np

data = pd.read_csv('projets.csv') ;
display( data )
```

Python

	projet	durée	durée cours	durée supplémentaire
0	Application santé	70h	25h	30h
1	Besoin conso batiments	100h	47h	40h
2	Segmentation clients e-commerce	70h	25h	20h
3	Catégorisation questions	80h	8h	80h
4	Classification images	100h	40h	60h
5	Preuve de concept	30h	1h	20h
6	Compétition Kaggle	50h	0h	50h

Notebook pour estimer les **dates prévisionnelles** de soutenances :

Vérification du type des colonnes et conversion des durées str->int

- Toutes les colonnes sont vérifiées avec une boucle for
- si la colonne est déjà en int, ou si 'durée' n'est pas présent dans la key, alors la colonne est inchangée

```
for key in data.keys():
    if (data[key].dtype == 'int64') or ( not 'durée' in key): continue ;
    data[key] = pd.to_numeric(data[key].str.replace('h',''), errors='coerce' ) ;
# for #
print('dtypes:\n', data.dtypes)
```

Python

```
dtypes:
  projet                object
  durée                 int64
  durée cours           int64
  durée supplémentaire  int64
dtype: object
```

Notebook pour estimer les **dates prévisionnelles** de soutenances :

Calcul de la durée totale pour chaque projet

```
data['durée totale'] = data['durée'] + data['durée cours'] + data['durée supplémentaire'] ;  
display( data )
```

Python

	projet	durée	durée cours	durée supplémentaire	durée totale
0	Application santé	70	25	30	125
1	Besoin conso batiments	100	47	40	187
2	Segmentation clients e-commerce	70	25	20	115
3	Catégorisation questions	80	8	80	168
4	Classification images	100	40	60	200
5	Preuve de concept	30	1	20	51
6	Compétition Kaggle	50	0	50	100

Notebook pour estimer les **dates prévisionnelles** de soutenances :

Temps initial estimatif en heures

1. calcul du nombre de jours de janvier à juin inclus
2. calcul du nombre de semaines (entières)
3. calcul du nombre d'heures à disposition sur une base de 35h/semaine
4. somme des heures des différents projets (estimation) et comparaison avec le nombre d'heures à disposition

```
n_jours = np.sum( [31,28,31,30,31,30] ) ;
n_sem = n_jours // 7 +2 ; # floor + ajout 2 semaines
n_jours_restant = n_jours - 7*n_sem ;
n_heures_dispo = 35 * n_sem ;
heures_total_estim = data['durée totale'].sum() ;
print( 'nombre initial de jours : {}'.format(n_jours) ) ;
print( 'nombre de semaines : {}'.format(n_sem) ) ;
print( 'nombre de jours restants : {}'.format(n_jours_restant) ) ;
print( "nombre d'heures disponibles : {}".format(n_heures_dispo) ) ;
print( "nombre total d'heures estimées : {}".format(heures_total_estim) ) ;
```

Python

```
nombre initial de jours : 181
nombre de semaines : 27
nombre de jours restants : -8
nombre d'heures disponibles :945
nombre total d'heures estimées : 946
```

Notebook pour estimer les **dates prévisionnelles** de soutenances :

Estimation de la durée de chaque projet

Valeurs en semaines, sur une base de 35h/semaine, arrondies à l'entier supérieur

```
key = 'durée estimée (semaines)' ;
data[key] = ( data['durée totale'] / 35 + 1. ).astype('int') ;
print('nombre initial de semaines (janvier à juin) : {}'.format( n_sem ) ) ;
print('nombre de semaines estimé : {}'.format( data[key].sum() ) ) ;
display( data )
```

Python

nombre initial de semaines (janvier à juin) : 27
nombre de semaines estimé : 30

	projet	durée	durée cours	durée supplémentaire	durée totale	durée estimée (semaines)
0	Application santé	70	25	30	125	4
1	Besoin conso batiments	100	47	40	187	6
2	Segmentation clients e-commerce	70	25	20	115	4
3	Catégorisation questions	80	8	80	168	5
4	Classification images	100	40	60	200	6
5	Preuve de concept	30	1	20	51	2
6	Compétition Kaggle	50	0	50	100	3

Notebook pour estimer les **dates prévisionnelles** de soutenances :

Estimation des dates de soutenance

1. Conversion des durées cumulées en *timedelta*
2. Ajout à une date initiale le 02/01/2023

```
duree_cumsum = 7*data['durée estimée (semaines)'].cumsum() ;
data['date soutenance (estimation)'] = \
    pd.to_datetime( '02/01/2023',format='%d/%m/%Y') \
    + pd.to_timedelta( duree_cumsum.astype(str) + ' days') ;
display(data)
```

Python

	projet	durée	durée cours	durée supplémentaire	durée totale	durée estimée (semaines)	date soutenance (estimation)
0	Application santé	70	25	30	125	4	2023-01-30
1	Besoin conso batiments	100	47	40	187	6	2023-03-13
2	Segmentation clients e-commerce	70	25	20	115	4	2023-04-10
3	Catégorisation questions	80	8	80	168	5	2023-05-15
4	Classification images	100	40	60	200	6	2023-06-26
5	Preuve de concept	30	1	20	51	2	2023-07-10
6	Compétition Kaggle	50	0	50	100	3	2023-07-31

Notebook pour estimer les **dates prévisionnelles** de soutenances :

Création de la chaine de caractères à copier/coller dans la fiche de début de formation

```
str_out = '' ;
for i, date in enumerate( data['date soutenance (estimation)'].tolist() ):
    date_str = date.strftime('%d/%m/%Y') ;
    str_out += '\n-Projet {}: {}'.format( i+1, date_str ) ;
# for i, date #
print(str_out) ;
```

Python

```
-Projet 1: 30/01/2023
-Projet 2: 13/03/2023
-Projet 3: 10/04/2023
-Projet 4: 15/05/2023
-Projet 5: 26/06/2023
-Projet 6: 10/07/2023
-Projet 7: 31/07/2023
```

Notebook pour estimer les **dates prévisionnelles** de soutenances :

Création de la chaine de caractères à copier/coller dans la fiche de début de formation

```
str_out = '' ;  
for i, date in enumerate( data['date soutenance (estimation)'].tolist() ):  
    date_str = date.strftime('%d/%m/%Y') ;  
    str_out += '\n-Projet {}: {}'.format( i+1, date_str ) ;  
# for i, date #  
print(str_out) ;
```

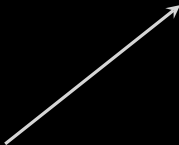
Python

```
-Projet 1: 30/01/2023  
-Projet 2: 13/03/2023  
-Projet 3: 10/04/2023  
-Projet 4: 15/05/2023  
-Projet 5: 26/06/2023  
-Projet 6: 10/07/2023  
-Projet 7: 31/07/2023
```

Date limite de fin de formation : 21/08/2023

Quelques bonnes pratiques





**Calendrier
(dates des jalons)**



Calendrier
(dates des jalons)

Technique POMODORO
(limite la fatigue)



Calendrier
(dates des jalons)

Technique POMODORO
(limite la fatigue)

Préparation
sessions mentorat



Calendrier
(dates des jalons)

Technique POMODORO
(limite la fatigue)

Préparation
sessions mentorat

Éviter de rester bloqué
(avancer sur autre chose)



Calendrier
(dates des jalons)



Technique POMODORO
(limite la fatigue)

Préparation
sessions mentorat

Éviter de rester bloqué
(avancer sur autre chose)



Calendrier
(dates des jalons)



Technique POMODORO
(limite la fatigue)



Préparation
sessions mentorat

Éviter de rester bloqué
(avancer sur autre chose)



Calendrier
(dates des jalons)



Technique POMODORO
(limite la fatigue)



Préparation
sessions mentorat



Éviter de rester bloqué
(avancer sur autre chose)



Calendrier
(dates des jalons)



Technique POMODORO
(limite la fatigue)



Préparation
sessions mentorat



Éviter de rester bloqué
(avancer sur autre chose)





Calendrier
(dates des jalons)



Technique POMODORO
(limite la fatigue)



Préparation
sessions mentorat



Éviter de rester bloqué
(avancer sur autre chose)



Outils collaboratifs

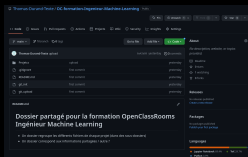
Outils de
communication
pour le **Mentorat**



Outils de communication pour le **Mentorat**



GitHub

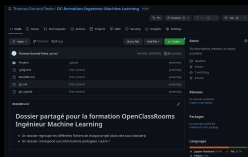


Outils de communication pour le **Mentorat**



GitHub

E-mails



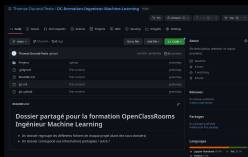
Outils de communication pour le **Mentorat**



GitHub

E-mails

OpenClassrooms



Merci pour votre attention.

Artworks

openclipart.org :

- loudspeaker by forestgreen
- Digital Marketing by GDJ
- Time Project by Ibear
- Laptop Computer Icon by kael_179
- Approved Business Stamp 1, Computer Handshake 1 by Merlin2525
- Glossy Smiley Set 3 by Chrisdesign
- mail-8 by gezegen

commons.wikimedia :

- Mass Spring System Resonance.gif by Guillermo Bossio

Divers:

- Letting Go & Leaning In - Turning Change Into Chance (pierretteraymond.com)
- dynamique-mag.com/article/etapes-cles-reussir-creation-entreprise.10007

Vecteezy:

- fond de tableau noir avec des éléments de fournitures scolaires by anuwat meerevee
- coach de vie pour la consultation, l'éducation, la motivation, la perspective de mentorat et l'auto-encadrement dans le modèle d'illustration plate de dessin animé dessiné à la main by denayunevz

Temporary page!

\LaTeX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because \LaTeX now knows how many pages to expect for this document.