



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

BILBOKO  
INGENIARITZA  
ESKOLA  
ESCUELA  
DE INGENIERÍA  
DE BILBAO

# ***APRENDIZAJE PROFUNDO POR REFUERZO PARA AUTÓMATAS CELULARES MULTI-POBLACIONALES: DISEÑO DE POLÍTICAS LOCALES, GRUPALES Y GLOBALES***

**Alumno:** Fernández López-Areal, Mateo

**Director:** Del Ser Lorente, Javier

GRADO EN INGENIERÍA EN TECNOLOGÍA DE TELECOMUNICACIÓN

**TRABAJO FIN DE GRADO**

# Índice

- Introducción
  - Objetivo
  - Contexto
- Estado del arte
  - Conceptos básicos
  - Alternativas
- Desarrollo
  - Creación del escenario
  - Metodología
  - Resultados
- Conclusión
  - Planificación
  - Coste

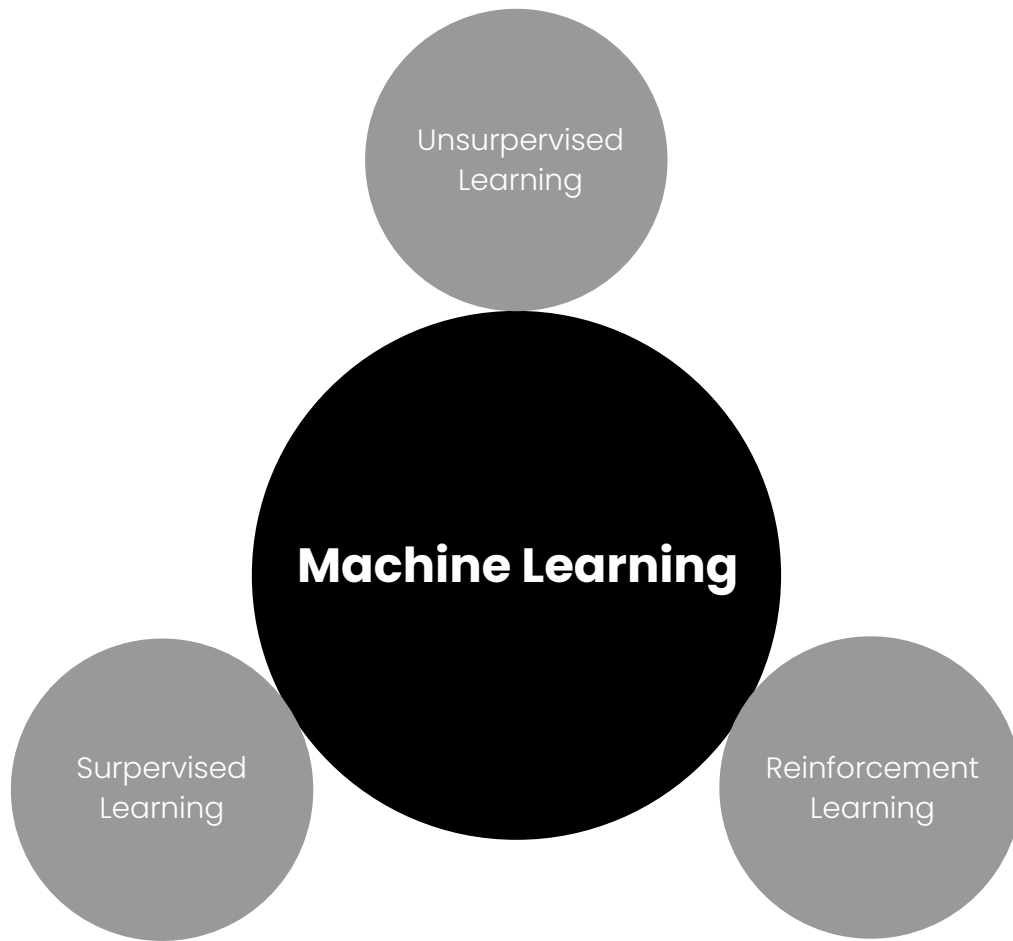


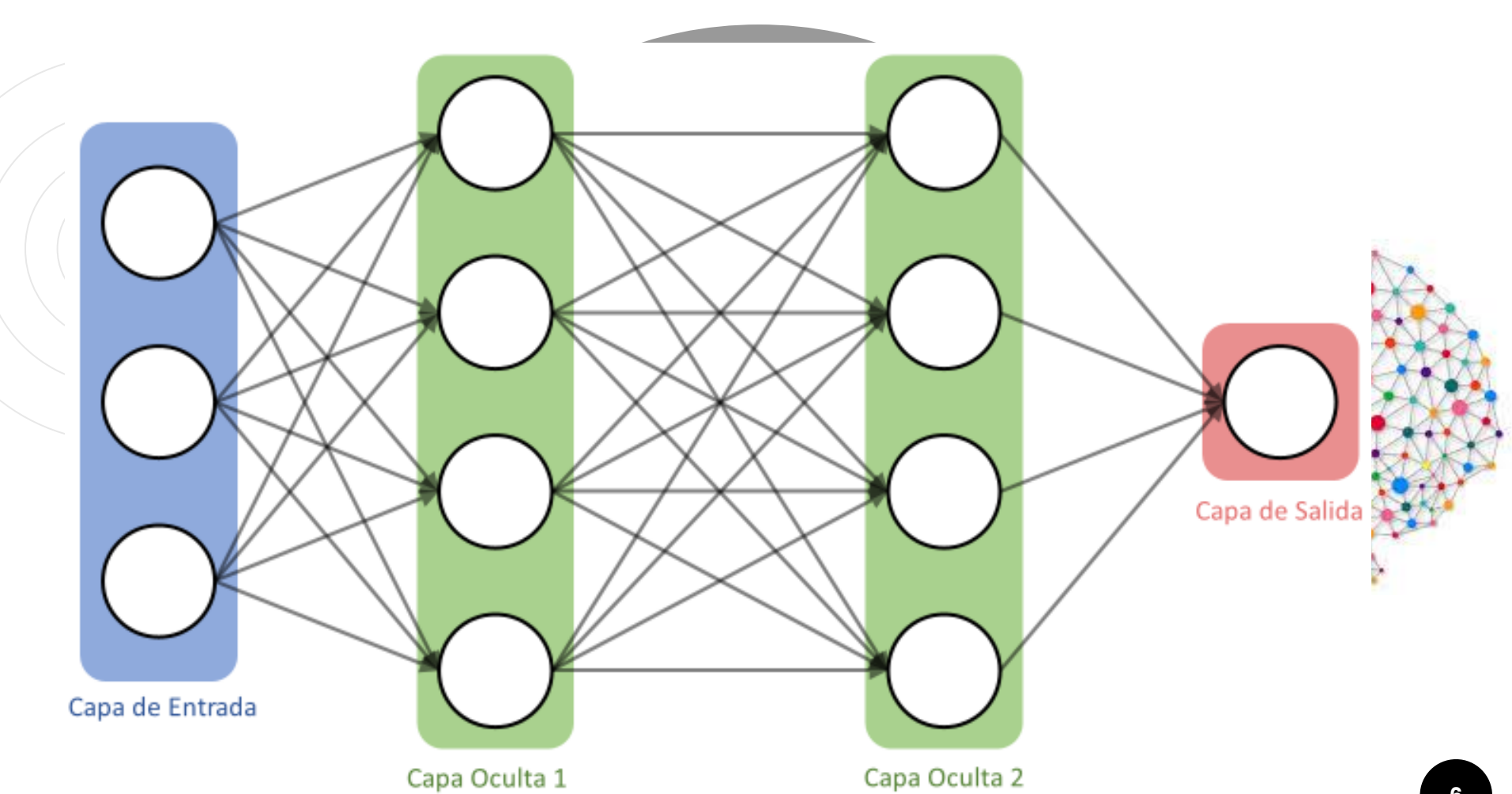
1

# Introducción



**Explorar cómo controlar  
las dinámicas  
comportamentales de  
sistemas multiagente  
multipoblacionales  
empleando aprendizaje  
por refuerzo profundo.**





# Contexto

- Necesidad de entender como opera la inteligencia artificial en diferentes escenarios
- Comprender cómo múltiples agentes interactúan entre sí en un entorno restringido





2

# **Estado del arte**



# Conceptos básicos

## Agente

Decide que acciones tomar sobre el entorno.

## Recompensa

Señal numérica que representa el grado de éxito de la acción

## Entorno

Todo con lo que pueda interactuar el agente de forma directa o indirecta

## Política

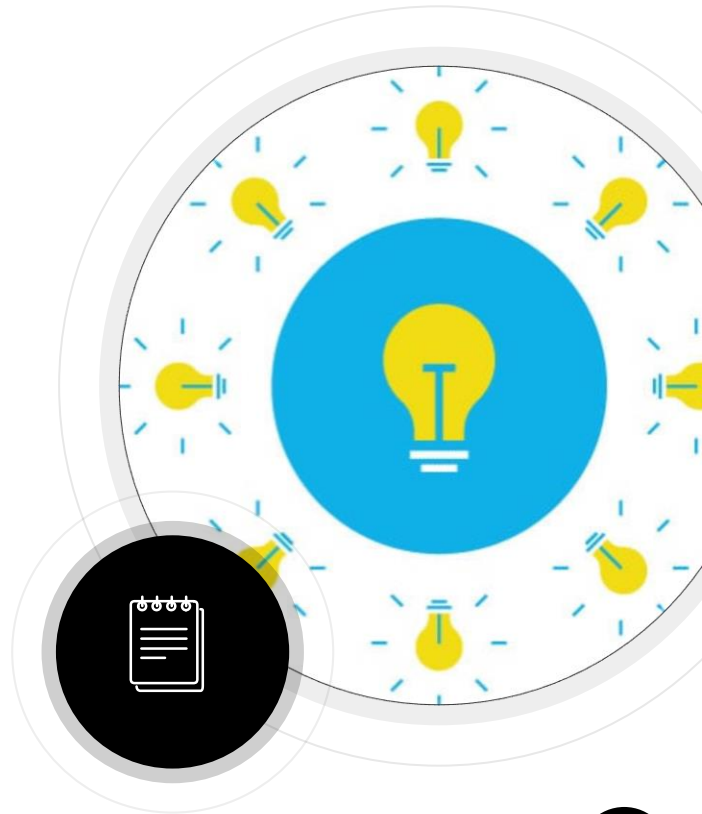
Es el mapeo estado-acción

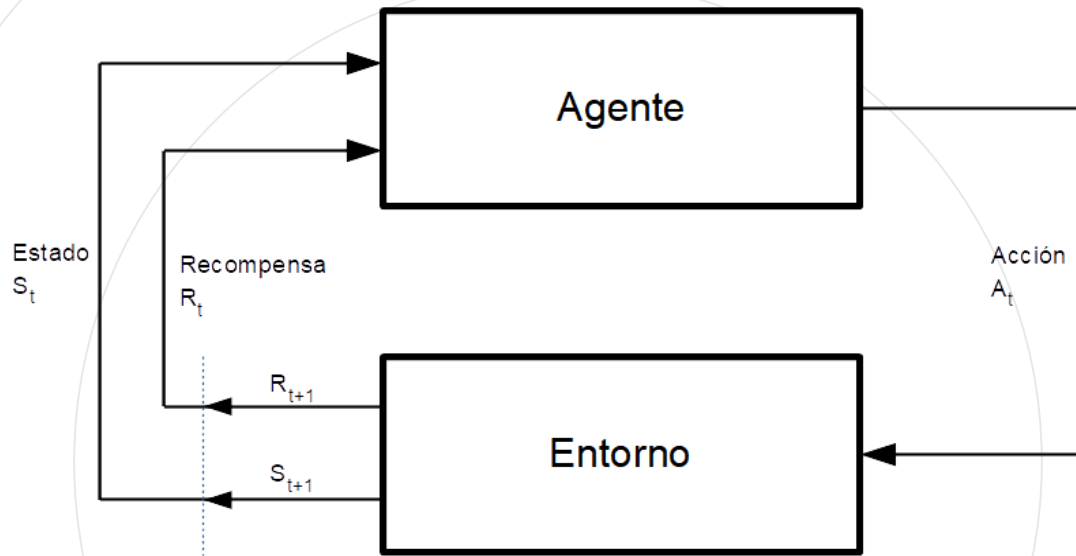
## Estado

Cada escenario en la que el agente se encuentra en el entorno

## Algoritmo de aprendizaje

Método en el que el modelo aprende





# Reinforcement Learning

Diagrama de bloques

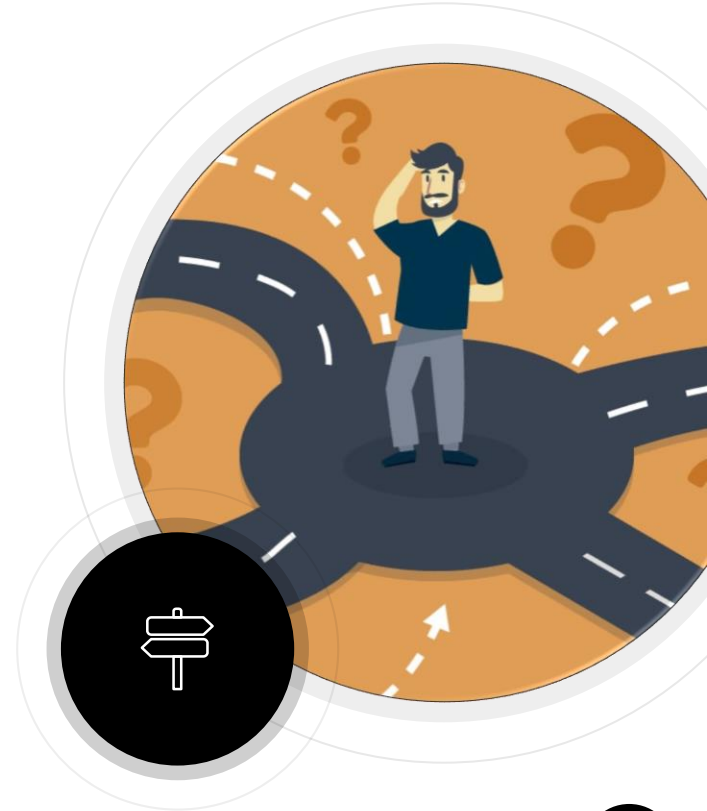
# Alternativas

## Algoritmos de aprendizaje

- Q-Learning
- Policy Gradient
- Actor-Critic

## Entornos de simulación

- AirSim
- ML-Agents
- DeepMind Lab
- OpenAI Gym



# **PPO + ML-Agents**



3

**Desarrollo**

# Creación del escenario

- Crear y programar los agentes
- Ejecutar el entrenamiento sobre el entorno de simulación
- Configurar los hiperparámetros

**Agentes** → Modelos 2D de cubos estáticos apilados en una matriz 12x12



# Metodología

## Local

Busca la supervivencia del agente

## Grupal

Busca que los agentes de una misma población se agrupen entre ellos en bloques con el fin de encontrar poblaciones diferenciadas en el entorno.

## Global

Busca que todas las poblaciones perduren en el tiempo, es decir, que no haya una sola población viva en el entorno.



# Resultados

## Estados del entorno

Observar como evoluciona el entorno sobre el simulador a cada paso

## Gráficas TensorBoard

Estudiar las gráficas de la recompensa acumulada y de la entropía obtenidas a través de TenorBoard.

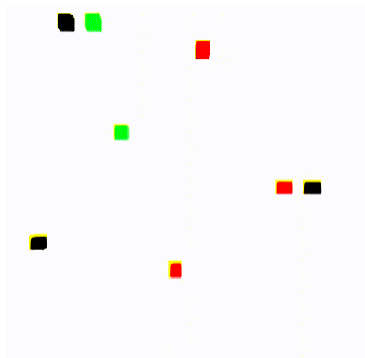




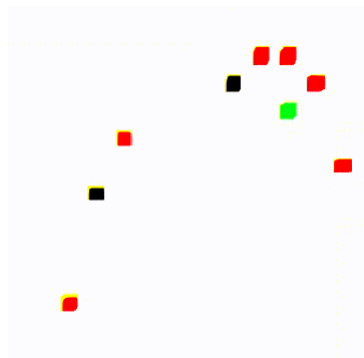
# Resultados

Local, grupal y global

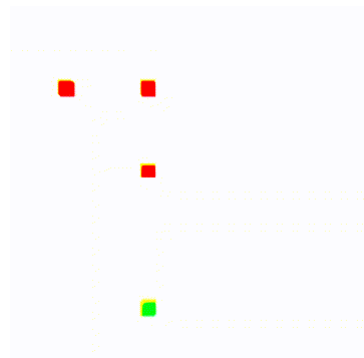
Local



Local y grupal



Vecinos



Grid





¿Beneficios?  
**Extrapolar**



4

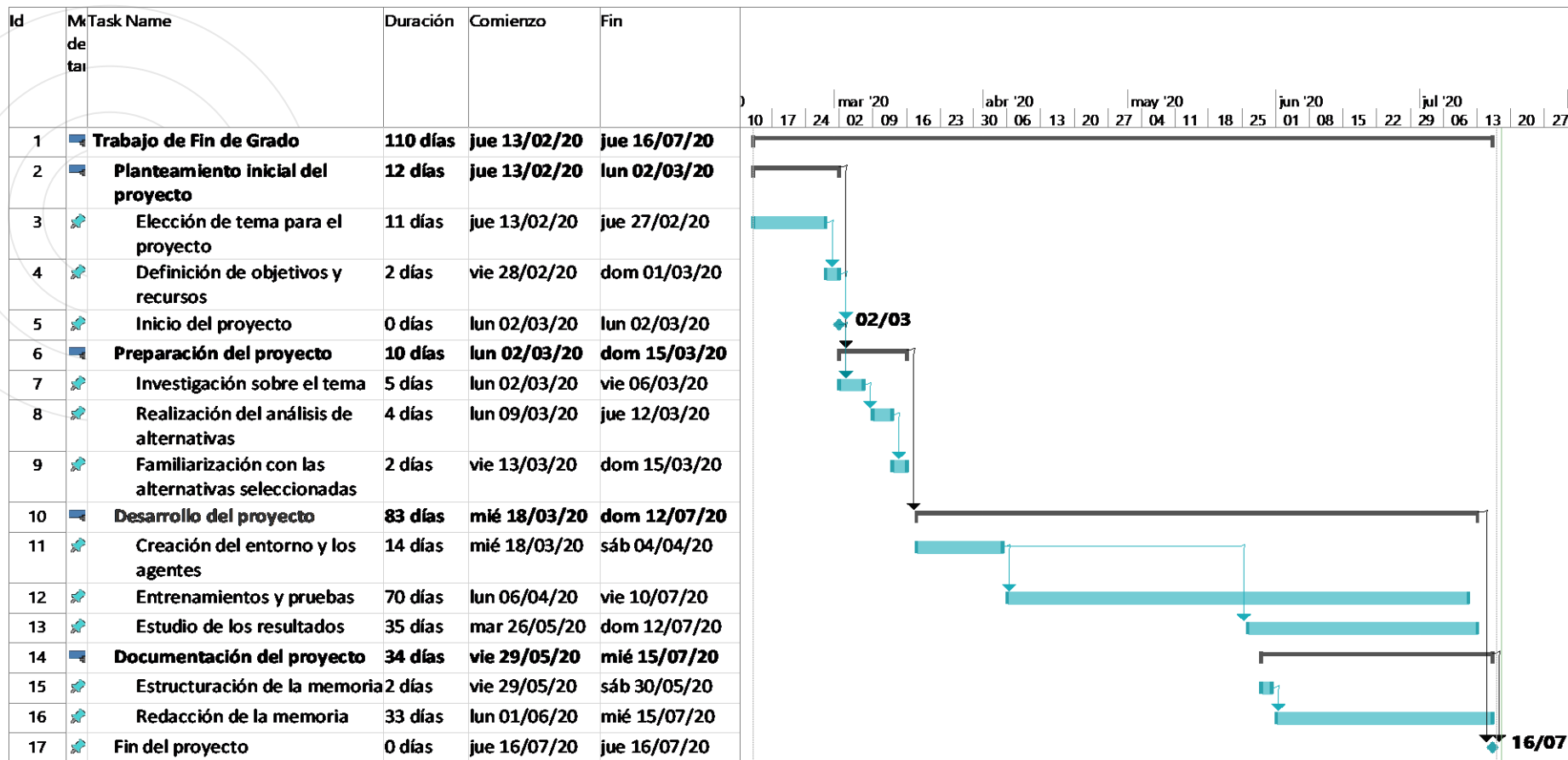
# Conclusión

# Planificación

Cuatro etapas diferenciadas:

- Planteamiento inicial del proyecto
- Preparación del proyecto
- Desarrollo del proyecto
- Documentación del proyecto







# Coste

Concepto	Coste (€)
Recursos humanos	14.500
Materiales amortizables	175
Documentación	90
<b>Subtotal</b>	<b>14.765</b>

# Conclusiones finales

- El agente siempre va a tender a evolucionar y sobrevivir.
- Cuanto más complejo se hace el sistema, más le cuesta al agente el aprendizaje.
- Si las soluciones técnicas empleadas son las adecuadas, el conocimiento adquirido por parte del agente tenderá a ser óptimo



# **¡Gracias!**

**¿Alguna duda?**