



# DelphiDay

italian conference

## AI – Un'introduzione tecnica

Tutte le novità su linguaggio e RTL nelle ultime versioni di Delphi



# LUCA MINUTI



[luca.minuti.it](https://luca.minuti.it)



[luca.minuti@gmail.com](mailto:luca.minuti@gmail.com)



[dev.to/lminuti](https://dev.to/lminuti)



[github.com/lminuti](https://github.com/lminuti)



[linkedin.com/in/luca.minuti/](https://linkedin.com/in/luca.minuti/)



**DelphiDay**  
italian conference

19 Novembre 2025  
Padova



**wintech**  
italia

# OPEN-SOURCE PROJECTS

[github.com/Iminuti](https://github.com/Iminuti)

## WiRL

[github.com/delphi-blocks/WiRL](https://github.com/delphi-blocks/WiRL)

## Delphi SAML

[github.com/EtheaDev/Delphi-SAML](https://github.com/EtheaDev/Delphi-SAML)

## OpenSSL

[github.com/Iminuti/Delphi-OpenSSL](https://github.com/Iminuti/Delphi-OpenSSL)

## MCPCConnect

[github.com/delphi-blocks/MCPCConnect](https://github.com/delphi-blocks/MCPCConnect)



19 Novembre 2025  
Padova





# AGENDA

---

1. Cenni storici
2. Cosa sono gli LLM
3. SmartCore AI Component Pack
4. Creare un chat bot
5. Modi alternativi per usare gli LLM

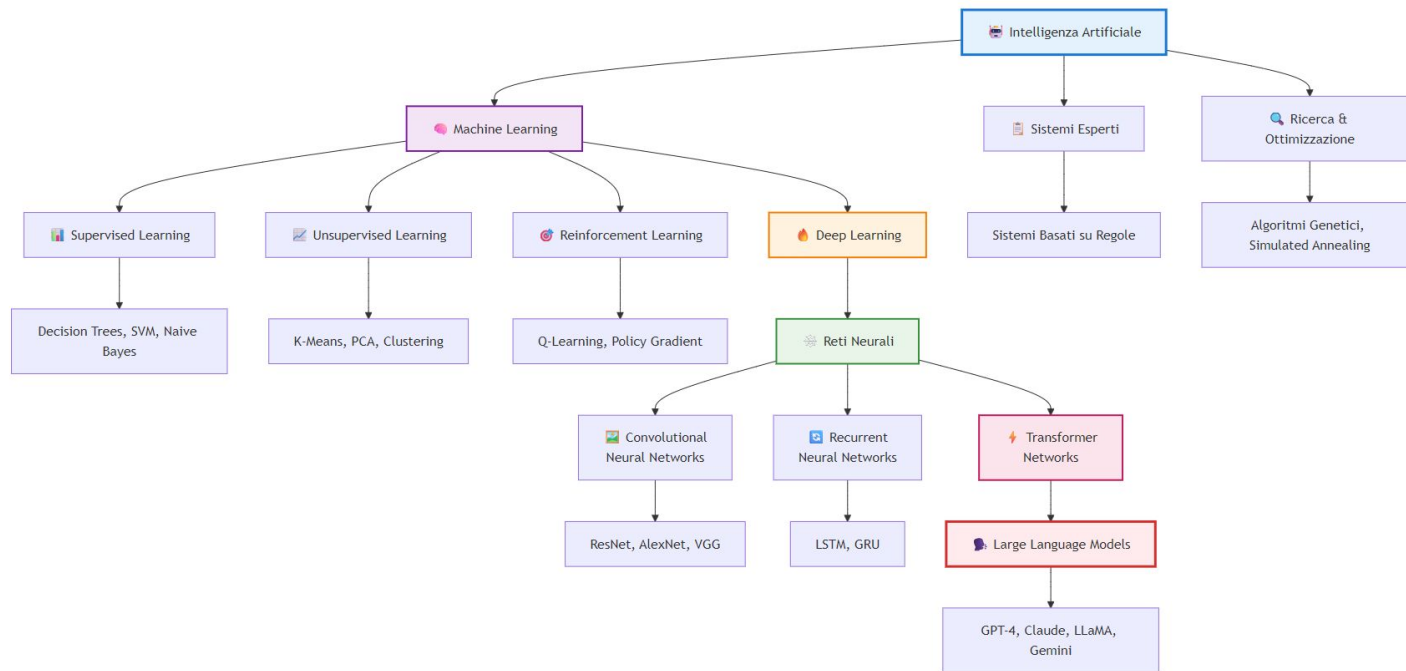


# INTRODUZIONE

- Il termine “Intelligenza Artificiale” è stato coniato da **John McCarthy** (creatore di LISP) nel 1956
- Ma già nel 1950 **Alan Turing** nell’articolo “Computing machinery and intelligence” proponeva il famoso **test di Turing** per valutare l’intelligenza di una macchina



# ALGORITMI





# INTRODUZIONE

- Anni 50-60: entusiasmo ma poco uso pratico
- Anni 70-80: “Inverno delle AI”, sistemi esperti
- Anni ‘90: Deep Blue, anti-spam, dati anziché regole
- Anni 2000-2010: Motori di ricerca, raccomandazioni
- Anni 2010-2020: traduzioni, riconoscimento facciale, assistenti vocali, guida autonoma





# INTRODUZIONE

- 2020-Oggi:
  - Modelli generativi (come GPT, DALL·E, Midjourney, Claude): capaci di scrivere testi, generare immagini, video, codice
  - ChatGPT porta l'AI conversazionale al grande pubblico
  - Integrazione in lavoro, creatività, sanità, educazione, marketing, automazione, supporto medico, assistenza clienti, progettazione, ricerca scientifica





# TRANSFORMER

- Paper "Attention Is All You Need": Introduce l'architettura Transformer (Google Brain 2017)
- Meccanismo di Self-Attention: Ogni token può "guardare" tutti gli altri simultaneamente
- Parallelizzazione: Training molto più veloce
- Encoder-Decoder: Struttura modulare e flessibile



# ADDESTRAMENTO

- **Pre-training:** Fase iniziale in cui il modello viene "allenato" su grandi quantità di testo o dati
- **Fine-tuning:** il modello viene "affinato" su compiti specifici con dataset annotati da esseri umani (es. domande/risposte, classificazioni, dialoghi, ecc.).
- **Reinforcement Learning (RLHF):** addestrato usando un meccanismo di rinforzo guidato da valutazioni umane



# TOKEN

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🍌🍌🍌🍌

Sequences of characters commonly found next to each other may be grouped together:  
1234567890

Text

Token IDs

<https://platform.openai.com/tokenizer>



# GLOSSARIO

- **Token:** Unità base di input/output per un modello di linguaggio. Può essere una parola, una sillaba o anche solo una lettera/punteggiatura.
- **Parametri:** I "pesi" interni del modello (milioni o miliardi)
- **Embedding:** Rappresentazione numerica (vettore) di parole, frasi o immagini
- **Multimodalità:** Capacità di elaborare più tipi di dati



DELPHI E LLM

1



# SMARTCORE AI

- Nuova libreria disponibile su **GetIt** per il supporto AI:
  - OpenAI, Gemini, Claude, Ollama
- Concetto simile ai database:
  - Driver, Connection, Chat/Image
- Risposte asincrone (tramite eventi)



# SMARTCORE AI

AI Connection Editor

AI Driver: Gemini

Parameter	Value	Default
BaseURL	<a href="https://generativelanguage.googleapis.com/v1beta/m">https://generativelanguage.googleapis.com/v1beta/m</a>	<a href="https://generativelanguage.googleapis.com/v1beta/m">https://generativelanguage.googleapis.com/v1beta/m</a>
APIKey		
Model	<a href="#">gemini-2.0-flash</a>	<i>gemini-2.0-flash</i>
MaxToken	1024	1024
Temperature	1	1
Timeout	60000	60000
TopP	1	1
TopK	1	1
SampleCount	1	1
AspectRatio		
SampleImageSize		
PersonGeneration	<a href="#">allow_adult</a>	<i>allow_adult</i>
Endpoint_GenerateContent	<a href="#">/%s:generateContent?key=%s</a>	<i>/ %s:generateContent?key=%s</i>
Endpoint_Models	<a href="#">?key=</a>	<i>?key=</i>

Test Connection OK Cancel





# SMARTCORE AI

---

```
procedure DemoForm.AIChatRequest1Response(Sender: TObject; const Text: string);  
begin  
    MemoResponse.Text := MemoResponse.Text + Text;  
end;  
  
procedure DemoForm.ChatButtonClick(Sender: TObject);  
begin  
    MemoResponse.Clear;  
    AIChatRequest1.Chat(MemoRequest.Text);  
end;
```



# SMARTCORE AI

- Perché avere una Chat nelle nostre APP?
  - RAG - retrieval augmented generation
  - ChatToSQL - Interrogare il database in linguaggio naturale
  - BlobToData - Dare un senso a dati testuali

demo time





# APPLICAZIONI

# 4



# APPLICAZIONI

- Prompt Augmentation: Arricchimento del contesto con dati proprietari (es. RAG)
- LLM as API: Funzionalità invisibili all'utente (estrazione di keyword, sintesi, ecc.)
- Natural Language to code: Query database in linguaggio naturale
- Agenti Autonomi: LLM che interagisce con tool e API



# RAG

- Il prompt utente viene arricchito con contesto rilevante
- Fonte dati: Database vettoriali, documenti, knowledge base
- Esempio: Chat aziendale con accesso a documentazione interna
- Vantaggio: Risposte accurate basate su dati aggiornati e specifici



# LLM AS API

- LLM come motore per operazioni dove manca API dedicata
- L'utente non percepisce la chiamata al LLM
- Casi d'uso:
  - Estrazione keyword da testo
  - Generazione automatica titoli/riassunti
  - Sentiment analysis
  - Categorizzazione contenuti





# NATURAL LANGUAGE TO SQL

- Query database in linguaggio naturale
- Processo:
  - Utente fa domanda ("Ordini di agosto?")
  - LLM riceve schema database
  - Genera query SQL
  - Ritorna risultati elaborati
- Vantaggio: Accesso dati senza conoscere SQL



# AGENTI AUTONOMI

- LLM che decide quali tool/API chiamare
- Function calling, MCP (Model Context Protocol)
- Esempi:
  - Prenotazioni automatiche
  - Compilazione assistita
  - Controlli e validazioni
- Combina ragionamento e azioni concrete

demo time





# TOOLS

# 4

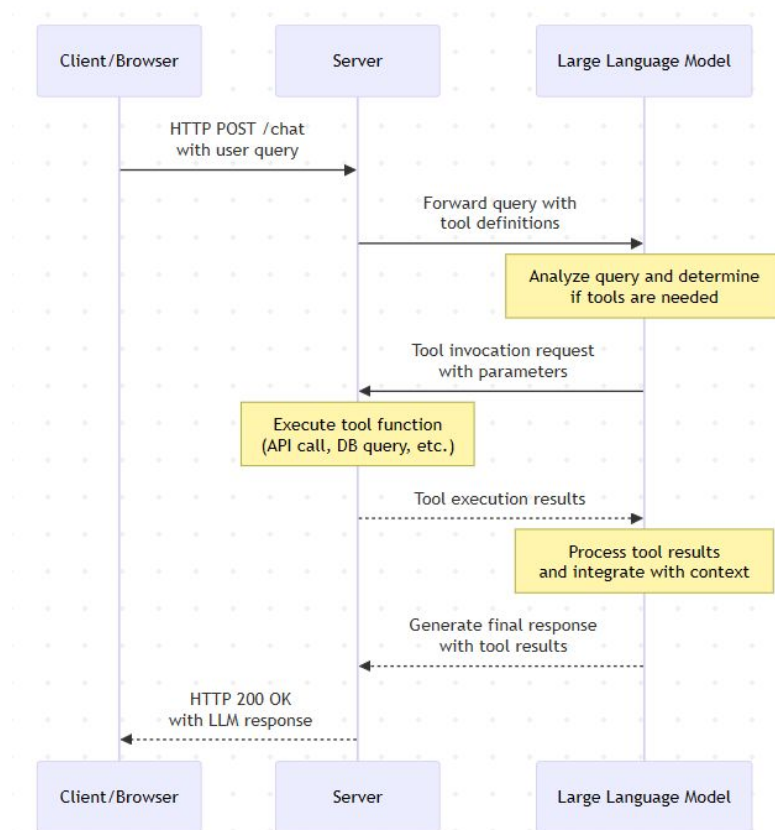


# TOOLS

- La maggior parte dei modelli permette di “allegare alla richiesta” una serie di tool
- Se il modello ritiene necessario l’uso dei tool per fornire una risposta migliore, invece di dare direttamente la risposta richiederà l’uso di un tool
- Solo dopo aver ottenuto i dati necessari produrrà la risposta



# TOOLS





THANK YOU