Introduzione al Natural Language Processing 🤖💬

Un viaggio nel mondo dell'elaborazione del linguaggio naturale

Benvenuti al corso che vi farà parlare con le macchine!

O almeno capire come fanno loro a parlare con noi...

"Alexa, cos'è l'NLP?" 🔊

> "L'NLP è la tecnologia che mi permette di capirti e risponderti. È il ponte tra il linguaggio umano e il mondo digitale."

Quante volte oggi avete già parlato con un'intelligenza artificiale? 🤔

Forse più di quante pensiate... dall'assistente vocale sul telefono al chatbot del servizio clienti, fino alla traduzione automatica di quel post in lingua straniera sui social media.

Fun fact: Alexa non sa ancora riconoscere il sarcasmo... quindi siate gentili con lei! 😉

Cosa tratteremo oggi 📋

1. **Cos'è il Natural Language Processing** 🧠

   Spoiler: non è insegnare a Siri le parolacce

2. **Importanza dell'NLP nel contesto aziendale** 💼

   Perché le aziende spendono miliardi su questa tecnologia?

3. **Problemi risolti dall'NLP** 🛠️

   Da "non capisco cosa dice il cliente" a "ho analizzato 1000 recensioni in 5s"

4. **Evoluzione storica dell'NLP** ⏳

   Un viaggio da "tradurre 7 frasi in russo" a "scrivere poesie come Shakespeare"

5. **Applicazioni emergenti dell'NLP** 🚀

   Il futuro è già qui, solo non è ancora uniformemente distribuito

NLP / AI

Cos'è il Natural Language Processing? 🧠

> "L'NLP è la disciplina che si occupa dell'interazione tra i computer e il linguaggio umano"

- Branca dell'intelligenza artificiale 🤖
- Intersezione tra:
  - 📚 Linguistica computazionale
  - 💻 Informatica
  - 📊 Machine learning

Un momento storico straordinario

L’intelligenza artificiale linguistica:

- Rivoluziona il modo di comunicare con le macchine
- Ridefinisce la programmazione
- Cambia la collaborazione tra sistemi
- Ridefinisce le basi dell’etica tecnologica

1️⃣ Vibe Coding

La Nuova Era della Programmazione Creativa

- Progetti digitali descritti a parole, senza scrivere codice
- Creatività e intuizione guidano lo sviluppo
- L'AI traduce le idee in soluzioni funzionanti

Nel corso:

- Accelerare la prototipazione
- Democratizzare l'innovazione
- Rendere la tecnologia più accessibile

2️⃣ Welfare dei Modelli

Etica e Responsabilità nell'AI

- Sistemi sempre più avanzati
- Domande su considerazione morale e segnali di "disagio"
- Rispetto, trasparenza e consapevolezza etica

Nel corso:

- Progettare soluzioni potenti e responsabili

3️⃣ Collaborazione tra Sistemi

MCP e A2A: L'AI che lavora in rete

- **Model Context Protocol (MCP)**
  Connette AI a dati, strumenti e repository esterni
- **Agent-to-Agent Protocol (A2A)**
  Permette agli agenti AI di comunicare e coordinarsi

Chiave per un'AI distribuita:
Agenti specializzati che collaborano per soluzioni modulari e sicure.

Concetti Chiave

- L’NLP unisce linguistica computazionale, informatica e machine learning
- Obiettivo: comprendere e generare linguaggio umano
- Va oltre la sintassi, comprende semantica e contesto

Il linguaggio naturale: una sfida complessa 🧩

Livello di complessità	Esempi di sfide
Sintattico	Struttura grammaticale, parsing delle frasi
Semantico	Significato delle parole e frasi, ambiguità
Pragmatico	Intenzioni, contesto, conoscenza del mondo
Culturale	Espressioni idiomatiche, riferimenti culturali

Importanza dell'NLP nel contesto aziendale 📊

La rivoluzione dei dati non strutturati 💾➡️📝

> "L'80% dei dati aziendali è non strutturato!" 📈

Cosa sono i dati non strutturati? Email, documenti, social media, recensioni, trascrizioni di chiamate, note... tutto ciò che non si adatta facilmente a righe e colonne di un database tradizionale.

- **Interagiscono** con i clienti 👥
- **Analizzano** enormi volumi di testo 📝
- **Ottimizzano** processi decisionali ⚙️
- **Scoprono** insights nascosti nei dati 💡

Domanda provocatoria: Pensate a tutti i dati testuali che la vostra organizzazione genera

Applicazioni aziendali dell'NLP 💼

- Analisi di mercato e intelligence competitiva 📈
- Ottimizzazione dei processi di R&D 🔬
- Miglioramento dell'esperienza cliente 😊
- Analisi dei feedback e voice of customer 👂

Case Study: NLP nel settore finanziario 💰

Le istituzioni finanziarie utilizzano l'NLP per:

- Analizzare report finanziari e notizie economiche
- Automatizzare compliance e rilevamento frodi
- Migliorare il servizio clienti con assistenti virtuali
- Valutare il sentiment degli investitori

> **JPMorgan Chase**: COIN (Contract Intelligence) analizza accordi di prestito in secondi invece di 360.000 ore/anno di lavoro umano!

Quiz! 🎯

Quali di questi problemi può risolvere l'NLP?

A) Classificazione automatica dei testi

B) Analisi delle immagini

C) Estrazione di informazioni da testi

D) Progettazione di circuiti elettronici

E) Analisi del sentiment

Problemi risolti dall'NLP 🛠️

L'NLP trasforma testi non strutturati in informazioni strutturate e actionable:

- **Classificazione automatica dei testi** 📄
- **Estrazione di informazioni** 🔍
- **Analisi del sentiment** 😊😐😠
- **Risposta automatica e chatbot** 💬
- **Riassunto automatico** 📝

Classificazione automatica dei testi

- Categorizza automaticamente documenti in base al contenuto
- **Settore sanitario**: classificazione di cartelle cliniche per patologie
- **Settore legale**: categorizzazione di contratti e sentenze

Estrazione di informazioni 🔍

- Identifica e estrae dati specifici da testi non strutturati
- **Settore immobiliare**: estrazione di prezzo, metratura, caratteristiche da annunci
- **Media**: estrazione di entità (persone, luoghi, organizzazioni) da articoli

Analisi del sentiment 😊😐😠

- Determina atteggiamento, opinione o emozioni in un testo
- **Marketing**: monitoraggio della percezione del brand
- **Entertainment**: valutazione della reazione del pubblico

Chatbot e sistemi di risposta automatica 💬

- Simulano conversazioni umane
- **E-commerce**: assistenza nella ricerca prodotti e acquisto
- **Educazione**: tutor virtuali personalizzati

Riassunto automatico 📝

- Condensa testi lunghi mantenendo le informazioni essenziali
- **Ricerca accademica**: generazione di abstract
- **Informazione**: creazione di riassunti di notizie da diverse fonti

Gli inizi: approcci basati su regole (1950-1980) 📚

- Esperimento Georgetown-IBM (1954): prima traduzione automatica
- Sistemi basati su regole linguistiche codificate manualmente
- ELIZA (1966): primo chatbot basato su pattern matching
- Limiti: ambiguità e complessità del linguaggio naturale

L'era statistica (1980-2010) 📊

- Rivoluzione: approcci basati su probabilità invece che regole
- Modelli n-gram e Markov nascosti
- Traduzione automatica statistica di IBM
- Apprendimento da grandi corpora di testo

La rivoluzione del deep learning (2010-presente) 🧠

- Word embeddings: Word2Vec (2013), GloVe (2014)
- Reti neurali ricorrenti (RNN, LSTM, GRU)
- 2017: Architettura Transformer ("Attention is All You Need")
- Modelli pre-addestrati: BERT, GPT
- Large Language Models (LLMs)

Timeline dell'evoluzione dell'NLP 📅

- **1954**: Esperimento Georgetown-IBM - Prima traduzione automatica
- **1966**: ELIZA - Primo chatbot basato su pattern matching
- **1980s**: Approcci statistici iniziano a sostituire i sistemi basati su regole
- **2013**: Word2Vec - Rivoluzione nella rappresentazione delle parole
- **2017**: "Attention is All You Need" - Introduzione dell'architettura Transformer
- **2018-2023**: Era dei modelli pre-addestrati e LLMs (BERT, GPT, ecc.)

Evuluzione degli LLM ⏳

https://arxiv.org/pdf/2303.18223

Sfide attuali nell'NLP 🧗

Anche i giganti hanno i loro talloni d'Achille!

Nonostante i progressi straordinari, l'NLP continua ad affrontare sfide significative che riflettono la complessità intrinseca del linguaggio umano:

- **Ambiguità linguistica** 🤔
  Quando "la vecchia porta la sbarra" può significare due cose completamente diverse
- **Multilingualità e localizzazione** 🌍
  Perché tradurre "sono al verde" in inglese letteralmente sarebbe un disastro
- **Comprensione del contesto e conoscenza del mondo** 🧠
  Capire che "ho preso un granchio" potrebbe non avere nulla a che fare con i crostacei

Sfide attuali nell'NLP 🧗

- **Robustezza e generalizzazione** 💪
  Sistemi che crollano se scrivi "gattto" invece di "gatto"
- **Efficienza computazionale e sostenibilità** ♻️
  Modelli che consumano energia come una piccola città

Ambiguità linguistica 🤔

Il mal di testa dei sistemi NLP!

- Ambiguità a vari livelli:
  - **Lessicale** (parole con più significati)
    "Amo la pesca" - Il frutto o l'attività sportiva? 🍑🎣
  - **Sintattico** (strutture frasali ambigue)
    "Ho visto la ragazza con il binocolo" - Chi aveva il binocolo?
  - **Semantico** (significati multipli di una frase)
    "Visiting professors can be boring" - I professori in visita possono annoiarsi o essere noiosi?
  - **Pragmatico** (intenzioni comunicative non esplicite)
    "Hai l'ora?" - Richiesta di informazione o richiesta di prestare l'orologio?

Ambiguità linguistica 🤔

> - "Ho visto l'uomo con il telescopio" 🔭
> - Ho usato un telescopio per vedere l'uomo?
> - Ho visto un uomo che aveva un telescopio?

Domanda interattiva: Potete pensare ad altri esempi di frasi ambigue nella vostra lingua? Come potrebbe un sistema NLP tentare di disambiguarle? 🧩

Multilingualità e localizzazione 🌍

Perché l'NLP non parla solo inglese!

- **Sviluppare sistemi NLP efficaci in tutte le lingue**
  L'inglese rappresenta solo il 25% di internet, ma riceve il 90% delle risorse di ricerca
- **Sfida particolare per lingue con risorse limitate**
  Delle 7.000+ lingue parlate nel mondo, solo circa 100 hanno risorse NLP sufficienti

Multilingualità e localizzazione 🌍

- **La localizzazione va oltre la traduzione**:
  - **Espressioni idiomatiche**: *"Costare un occhio della testa" ≠ "Cost an eye of the head"* 👁️
  - **Riferimenti culturali**: *Capire cosa significa "fare il Fantozzi" in Italia* 🇮🇹
  - **Norme comunicative**: *In giapponese, i livelli di formalità cambiano completamente la struttura della frase* 🇯🇵

Domanda di riflessione: In che modo le differenze culturali possono influenzare l'efficacia di un sistema NLP quando viene applicato in contesti linguistici diversi? 🌐

Comprensione del contesto e conoscenza del mondo

I modelli NLP spesso mancano di:

- **Comprensione del contesto più ampio**
  Capire che "è freddo" significa cose diverse in estate o in inverno
- **Conoscenza del mondo data per scontata dagli umani**
  Sapere che le persone non possono volare, che l'acqua è bagnata, che il sole sorge a est...
- **Capacità di fare inferenze basate su common sense**
  "Ho lasciato il gelato fuori dal freezer" → "probabilmente si è sciolto"

> "Non posso accedere al mio account dopo l'aggiornamento di ieri"
> Un sistema NLP dovrebbe capire che "l'aggiornamento" si riferisce probabilmente a un software e potrebbe essere la causa del problema.

Comprensione del contesto e conoscenza del mondo

Esempio divertente: Se dico "Ho preso un caffè con mia sorella, era davvero amaro", un umano capisce che era amaro il caffè, non mia sorella! 😅

Domanda tecnica: Quali approcci potrebbero aiutare i sistemi NLP a sviluppare una migliore comprensione del contesto e della conoscenza del mondo? 🌐

Robustezza e generalizzazione 💪

- I modelli NLP possono degradarsi significativamente con:
  - Input fuori distribuzione
  - Errori ortografici
  - Gergo specialistico
  - Linguaggio non standard
- Vulnerabilità ad attacchi adversarial

Efficienza computazionale e sostenibilità ♻️

- I modelli NLP più avanzati richiedono risorse enormi:
  - Addestramento di GPT-3: energia equivalente a centinaia di case americane in un anno
- Questioni di:
  - Accessibilità (democratizzazione della tecnologia)
  - Sostenibilità ambientale
  - Implementazione su dispositivi edge

Applicazioni emergenti dell'NLP 🚀

Il futuro è già qui, solo non è ancora uniformemente distribuito!

Nuove ed entusiasmanti direzioni di sviluppo stanno ridefinendo i confini dell'NLP:

- **NLP multimodale** 🖼️🔊
  Sistemi che vedono, ascoltano e parlano: "Descrivi questa immagine e trasforma la descrizione in una canzone"
- **Agenti conversazionali avanzati** 🤖
  Da semplici chatbot a veri assistenti che ricordano le conversazioni passate e si adattano alle tue preferenze
- **NLP per lingue a basse risorse** 🌏
  Democratizzare l'NLP per le 7.000+ lingue del mondo, non solo per le 10 più parlate

Applicazioni emergenti dell'NLP 🚀

Il futuro è già qui, solo non è ancora uniformemente distribuito!

Nuove ed entusiasmanti direzioni di sviluppo stanno ridefinendo i confini dell'NLP:

- **NLP interpretabile e spiegabile** 🔍
  Sistemi che non solo danno risposte, ma spiegano anche il perché: "Ho consigliato questo prodotto perché..."
- **NLP per il bene sociale** ❤️
  Tecnologie linguistiche per affrontare sfide globali: dalla disinformazione alle crisi umanitarie

Domanda visionaria: Quali di queste applicazioni emergenti pensate possa avere il maggiore impatto sulla società nei prossimi 5-10 anni e perché? 🌠

NLP multimodale 🖼️🔊

- Integra l'elaborazione del linguaggio con altre modalità:
  - Immagini
  - Video
  - Audio
- **Retail**: analisi combinata di recensioni testuali e immagini dei prodotti
- **Sanità**: correlazione tra sintomi descritti e pattern visibili nelle immagini mediche

Agenti conversazionali avanzati 🤖

- Evoluzione da semplici chatbot a sistemi sofisticati:
  - Conversazioni estese e contestuali
  - Comprensione di intenzioni complesse
  - Memoria delle interazioni precedenti
  - Adattamento al feedback dell'utente
- **Educazione**: tutor personalizzati
- **Salute mentale**: supporto iniziale e gestione dello stress

NLP per lingue a basse risorse 🌏

- Tecniche per lingue con pochi dati di addestramento:
  - Transfer learning multilingue
  - Few-shot learning
  - Generazione di dati sintetici
- **Conservazione culturale**: preservazione di lingue minoritarie
- **Settore umanitario**: comunicazione durante crisi in regioni con lingue locali

NLP interpretabile e spiegabile 🔍

- Modelli capaci di fornire giustificazioni comprensibili per le loro previsioni
- **Settore legale**: riferimenti specifici a precedenti o normative
- **Settore finanziario**: spiegabilità per valutazione del rischio e identificazione frodi

NLP per il bene sociale ❤️

- Applicazioni dell'NLP per affrontare sfide sociali:
  - Mitigazione della disinformazione
  - Risposta alle crisi
  - Monitoraggio ambientale
  - Accessibilità per persone con disabilità

Case Study: NLP nella ricerca farmaceutica 💊

Le aziende farmaceutiche utilizzano l'NLP per accelerare la scoperta di farmaci:

- Analisi automatica di milioni di articoli scientifici e brevetti
- Identificazione di relazioni tra geni, proteine, malattie e composti
- Estrazione di risultati da studi clinici
- Generazione di ipotesi per nuovi target terapeutici

> **BenevolentAI** ha identificato il baricitinib come potenziale trattamento per COVID-19 usando NLP, poi confermato da studi clinici!

Pensate a questo... 🤔

> "Il linguaggio è la casa dell'essere."
>
> - Martin Heidegger

Se il linguaggio definisce la nostra esperienza umana...
cosa significa creare macchine che lo comprendono?

Conclusione 🏁

- L'NLP è una frontiera in rapida evoluzione dell'IA
- Trasforma il modo in cui interagiamo con le macchine e le organizzazioni operano
- Sfide significative rimangono, ma nuove applicazioni continuano a emergere
- Comprendere fondamenti, capacità e limiti è essenziale per sfruttarne il potenziale

> "Il linguaggio è il software della mente." - Steven Pinker

Riferimenti e approfondimenti 📚

- Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed. draft). https://web.stanford.edu/~jurafsky/slp3/
- Vaswani, A., et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems, 30.
- Devlin, J., et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019.

Domande? 🙋 🙋‍♂️