CORSO IA CON CLAUDE · DISPENSA PER L'ALLIEVO

Ecosistema Claude, Constitutional AI e modelli

Chi costruisce lo strumento, come si comporta e quale versione usare

ORA	DURATA	MODULO
03 di 12	60 minuti	2 — Architettura

Gianluca Bonomo · www.gianlu.it · info@gianlu.it

Indice

Indice 2
Come usare questa dispensa 3
1. Chi è Anthropic 4
1.1 Un'azienda nata con un vincolo nello statuto 4
1.2 Constitutional AI: il modello che si corregge 4
1.3 Che cosa cambia, nell'uso di tutti i giorni 5
1.4 Modello, prodotto, interfaccia: tre cose diverse 7
2. La famiglia dei modelli 7
2.1 Tre tagli dello stesso strumento 8
2.2 Come si sceglie, in pratica 9
2.3 Quanto costa, e come si ragiona sul costo 9
2.4 I nomi delle versioni invecchiano in fretta 9
3. La finestra di contesto 10
3.1 Contesto grande non significa memoria 11
3.2 Quanto è, in pratica, un milione di token 11
3.3 Lavorare bene su una conversazione lunga 12
Le domande che arrivano in aula 13
3.4 Tre modi di dare il materiale, e quando usarli 13
4. Esercitazione pratica dell'ora 14
Glossario 16
Per approfondire 17

Come usare questa dispensa

Questa dispensa raccoglie, in forma scritta e ordinata, tutto quello che vediamo insieme durante l'ora di lezione. Non devi studiarla a memoria: serve a farti ritrovare le cose con calma, dopo, quando l'aula è finita e ti resta una domanda in sospeso.

È pensata per chi parte da zero. Ogni termine tecnico viene spiegato la prima volta che compare, e a fine dispensa trovi un glossario con tutte le parole nuove. Se una pagina ti sembra difficile, saltala e torna indietro dopo: nessun concetto di questa dispensa richiede conoscenze informatiche precedenti.

> **COME SONO FATTI I RIQUADRI**
>
> **In parole semplici** (azzurro): la spiegazione senza tecnicismi, quella da portarsi a casa se si ricorda una cosa sola.
>
> **Attenzione** (oro): un errore frequente, un rischio, una cosa da non fare.
>
> **Esempio** (grigio): un caso concreto, preso dalla vita di tutti i giorni o dal lavoro in azienda.

Obiettivo dell'ora

Capire chi è Anthropic e con quali priorità lavora, che cosa significa Constitutional AI, come è fatta la famiglia dei modelli Claude e che cosa cambia davvero avere una finestra di contesto molto ampia.

Che cosa saprai fare alla fine

- Spiegare a un collega la differenza fra i tre gradini della famiglia Claude e scegliere quello giusto per un compito dato.
- Capire perché un modello può rifiutare una richiesta, e che cosa significa che i criteri sono scritti.
- Stimare quanto materiale si può dare in pasto a una singola richiesta.
- Riconoscere quando conviene risparmiare e quando no.

1. Chi è Anthropic

Prima di affidare a uno strumento i documenti dell'azienda conviene sapere chi lo costruisce, con quali vincoli e con quali priorità.

1.1 Un'azienda nata con un vincolo nello statuto

Anthropic nasce nel **2021**, fondata da **Dario e Daniela Amodei** insieme ad altri ricercatori che venivano da OpenAI. La missione dichiarata è costruire sistemi di intelligenza artificiale sicuri e affidabili.

Fin qui, nulla che non direbbe qualsiasi azienda del settore. La differenza sta nella forma giuridica: Anthropic è una **Public Benefit Corporation**.

> **CHE COS'È UNA PUBLIC BENEFIT CORPORATION**
>
> È una società di capitali che, per statuto, deve perseguire **anche** una finalità di beneficio pubblico dichiarata, e non soltanto il profitto degli azionisti.
>
> Gli amministratori sono tenuti a bilanciare le due cose, e ne rispondono. Non è una dichiarazione d'intenti sul sito: è scritta nell'atto costitutivo.

Per chi valuta un fornitore questo è un dato di governance, non un dettaglio di colore: dice qualcosa su quali priorità l'azienda è obbligata a considerare quando le due cose entrano in conflitto.

1.2 Constitutional AI: il modello che si corregge

Nella terza ora del corso abbiamo visto che i modelli si allineano alle aspettative umane con tecniche come l'RLHF, l'apprendimento per rinforzo da feedback umano: delle persone valutano le risposte, il modello impara a produrne di più gradite.

Funziona, ma ha due limiti: è lento e non è trasparente. Chi ha deciso che quella risposta era migliore? Con quale criterio? Non è scritto da nessuna parte.

La proposta di Anthropic è un approccio diverso, chiamato **Constitutional AI**: dare al modello una "costituzione", cioè un insieme di principi espliciti e scritti, ispirati anche alla Dichiarazione universale dei diritti umani, e addestrarlo a usarli su sé stesso.



Il ciclo: il modello risponde, si critica sui principi, riscrive.

PERCHÉ CI INTERESSA IN AZIENDA

Il vantaggio pratico non è filosofico: è che i criteri sono **leggibili**.

Quando un modello rifiuta una richiesta o smussa una risposta, si può andare a vedere in base a quale principio lo ha fatto. E si può discuterne.

Con un criterio implicito, invece, resta solo da subire la decisione.

Nella pratica quotidiana questo si traduce in un modello che tende a essere prudente, a segnalare quando non sa, e a non assecondare richieste problematiche. Talvolta è più cauto del necessario: è il rovescio della stessa medaglia.

1.3 Che cosa cambia, nell'uso di tutti i giorni

Un modello addestrato così tende a comportarsi in tre modi riconoscibili, e conviene saperlo in anticipo per non prenderla come un difetto.

- **Segnala quando non sa.** Invece di riempire il vuoto con una risposta plausibile, tende a dire che l'informazione non c'è. È un pregio, anche se sul momento sembra una mancanza.
- **Smussa le richieste ambigue.** Se una domanda si presta a un uso problematico, spesso risponde alla versione innocua e lo dice.

- **Talvolta è più prudente del necessario.** Capita che rifiuti una richiesta legittima che assomiglia a una problematica: è il rovescio della stessa medaglia.

> **CHE FARE QUANDO RIFIUTA E HA TORTO**
>
> Quasi sempre il problema è il **contesto mancante**. Una richiesta come «scrivi un messaggio per convincere una persona a firmare» suona manipolatoria a freddo.
>
> Riscritta con il contesto — «sono il responsabile acquisti, devo sollecitare un fornitore che non ha ancora restituito il contratto firmato» — diventa una normale richiesta di lavoro.

Che cosa succede quando rifiuta

Prima o poi capita a tutti: una richiesta legittima viene rifiutata, o viene accolta con un preambolo di cautele che non c'entrano nulla. Succede perché il modello sta applicando i suoi principi a un caso che non ha inquadrato bene, non perché ci sia un divieto sul vostro lavoro.

Nella grande maggioranza dei casi la causa è il **contesto mancante**, e la soluzione è darglielo.

La richiesta che si blocca	La stessa richiesta che funziona
«Scrivimi una mail per contestare il comportamento di un dipendente»	«Sono il responsabile del personale. Devo scrivere una contestazione disciplinare formale, nei termini previsti dal contratto collettivo. Ecco i fatti…»
«Come si aggira questo controllo?»	«Sto verificando la sicurezza di un sistema di cui sono responsabile. Quali sono le debolezze tipiche di questo tipo di controllo, e come si correggono?»
«Dimmi che cosa ha questa persona»	«Questa è una descrizione di sintomi presa da un forum. Quali domande avrebbe senso porre a un medico?»

> **LA REGOLA PRATICA**
>
> Se una richiesta legittima viene rifiutata, **non riformulatela per aggirare il rifiuto**: spiegate chi siete e a che cosa vi serve.
>
> È anche un buon promemoria generale: un contesto scritto bene risolve molti più problemi di quanto sembri.

1.4 Modello, prodotto, interfaccia: tre cose diverse

In aula si fa quasi sempre confusione fra tre parole che sembrano sinonimi e non lo sono. Chiarirle adesso evita malintesi per tutto il resto del corso.

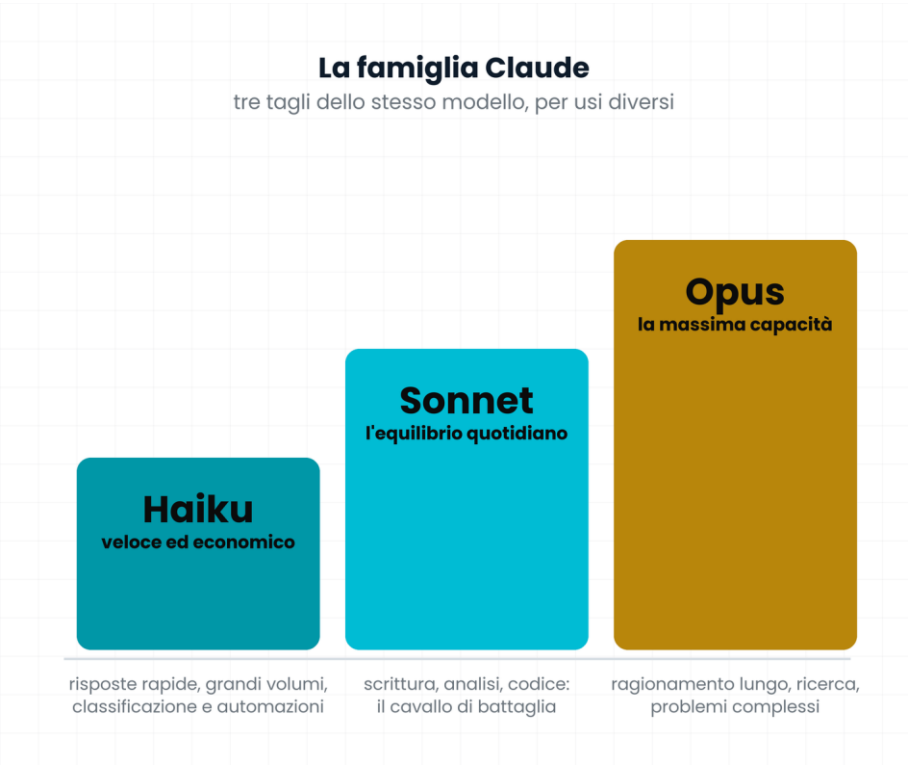
Il termine	Che cos'è	Quando ve ne accorgete
Il **modello**	Il motore: quello che è stato addestrato e che produce le risposte. Ha un nome e una versione	Quando scegliete fra il taglio rapido e quello potente
Il **prodotto**	Il servizio che vendete o comprate: il piano gratuito, quello individuale, quello aziendale, con le sue condizioni contrattuali	Quando si parla di dati, contratti e responsabilità: è tutta qui la quarta ora
L'**interfaccia**	Il modo in cui ci si arriva: la pagina web, l'applicazione per computer, l'integrazione in un altro programma	Quando qualcuno dice «da me si vede diverso»

Lo stesso modello si usa attraverso prodotti diversi e con garanzie contrattuali diverse. È il motivo per cui la domanda «è sicuro?» non ha una risposta unica: dipende dal prodotto, non dal modello.

2. La famiglia dei modelli

Non esiste “il modello migliore”. Esiste quello adatto al compito che hai davanti, e sceglierlo bene è una competenza.

2.1 Tre tagli dello stesso strumento



I tre gradini: cambiano i nomi delle versioni, non la logica.

Gradino	Come si comporta	Quando sceglierlo
Haiku il leggero	Molto rapido, costo basso, ottimo su compiti definiti.	Classificare, estrarre dati da elenchi, rispondere a domande semplici, gestire grandi volumi, automazioni.
Sonnet l'equilibrato	Buon ragionamento, buona velocità: il compromesso più usato.	Scrittura, analisi di documenti, codice, sintesi: copre la gran parte del lavoro d'ufficio.
Opus il potente	La massima capacità di ragionamento, più lento e più costoso.	Problemi a più passaggi, ricerca, analisi legali o finanziarie, casi in cui l'errore costa caro.

I NOMI INVECCHIANO, I GRADINI NO

Nel materiale del corso trovi nomi di versione precisi. Cambieranno: nel giro di un anno escono generazioni nuove e i numeri si spostano.

Quello che non cambia è la struttura a tre gradini e il criterio per sceglierli. Impara quello, e i nomi li leggerai di volta in volta dal menu.

2.2 Come si sceglie, in pratica

La domanda giusta non è «qual è il più bravo», ma **quanto costa sbagliare**.

- Se un errore si vede subito e si corregge in dieci secondi, usa il modello leggero: è più veloce e costa meno.
- Se l'errore finisce in un contratto, in un preventivo o in una risposta a un cliente, usa quello potente e **verifica comunque**.
- Se stai costruendo qualcosa che verrà riusato cento volte, vale la pena spendere di più una volta per farlo bene.

> **CINQUE CASI, CINQUE SCELTE**
>
> **Classificare 5.000 email per reparto** → leggero: compito ripetitivo e ben definito, l'errore si corregge.
>
> **Rispondere a un reclamo importante** → potente: c'è di mezzo un cliente e un tono da tenere.
>
> **Riassumere il verbale di una riunione** → equilibrato: basta e avanza.
>
> **Analizzare un contratto prima della firma** → potente, e rilettura umana.
>
> **Estrarre indirizzi da un elenco** → leggero.

2.3 Quanto costa, e come si ragiona sul costo

Il consumo si misura in **token**, sia in entrata sia in uscita. Come ordine di grandezza: mille token sono circa 750 parole di italiano, quindi una pagina e mezza di testo fitto.

Nei piani a sottoscrizione non si paga a token: si paga un canone e si ha un limite d'uso. Il ragionamento sul costo, quindi, per la maggior parte degli utenti aziendali si traduce in una domanda diversa: **quanto in fretta consumo il mio limite?**

> **LA REGOLA PRATICA**
>
> Un modello potente su un documento lungo consuma molto di più di un modello leggero su una domanda breve. Se vi capita di esaurire il limite a metà giornata, quasi sempre la causa è una sola: state usando il modello più potente anche per le cose banali.

2.4 I nomi delle versioni invecchiano in fretta

Ogni pochi mesi esce una versione nuova, con un numero diverso e capacità maggiori. È la parte del corso che invecchia più in fretta, e conviene saperlo fin da subito.

Quello che **non** cambia è la logica con cui si sceglie: tre tagli, uno economico e rapido, uno equilibrato, uno per i compiti difficili. Cambiano i nomi, non la domanda da farsi.

COME LEGGERE UNA NOTIZIA SU UN MODELLO NUOVO

Le tre cose da guardare sono sempre le stesse: **quanto contesto** regge, **quanto costa** rispetto alla versione precedente, **che cosa fa meglio** in concreto.

Tutto il resto — punteggi su prove di laboratorio, classifiche, dichiarazioni — dice poco su come si comporterà sul vostro lavoro.

Il modo affidabile di scegliere resta uno solo: provare la versione nuova su un compito vostro di cui conoscete già la risposta giusta.

3. La finestra di contesto

È il numero che più di ogni altro ha cambiato il modo di lavorare con questi strumenti.

La **finestra di contesto** è quanto testo il modello riesce a tenere sotto gli occhi in una sola richiesta: la tua domanda, i documenti allegati e la conversazione precedente, tutto insieme. Si misura in token, i pezzetti di testo di cui abbiamo parlato nella prima ora.



Ordini di grandezza: da un'email a un archivio documentale.

La generazione attuale dei modelli principali arriva a **un milione di token**; i modelli più leggeri si fermano intorno ai **200.000**, che sono già circa cinquecento pagine, cioè un libro intero.

CHE COSA CAMBIA DAVVERO

Fino a poco tempo fa bisognava **riassumere prima di chiedere**: spezzare un documento, darne dei pezzi, ricucire le risposte.

Oggi si fa il contrario: si dà **tutto** il materiale e si chiede dopo. Il modello vede le relazioni fra un capitolo e l'altro, cosa che con i pezzetti separati era impossibile.

CONTESTO GRANDE NON SIGNIFICA MEMORIA

La finestra vale per **quella** richiesta. Chiusa la conversazione, il contenuto non resta da nessuna parte.

Per avere documenti sempre disponibili servono i progetti, che vedremo nella nona ora.

3.1 Contesto grande non significa memoria

È la confusione più frequente. La finestra di contesto vale per **quella** richiesta: è come il tavolo su cui il modello lavora, non come un archivio.

Chiusa la conversazione, il contenuto non resta da nessuna parte: la volta dopo si ricomincia da zero. Per avere documenti sempre disponibili servono i **progetti**, che vedremo nella nona ora.

UNA CONSEGUENZA PRATICA POCO INTUITIVA

Anche dentro la stessa conversazione, più la chat si allunga più il materiale iniziale si allontana. Su conversazioni molto lunghe può capitare che il modello perda di vista un dettaglio dato all'inizio.

Se sta succedendo, la soluzione non è insistere: è ricominciare una conversazione nuova rimettendo il materiale essenziale.

3.2 Quanto è, in pratica, un milione di token

I numeri sulla finestra di contesto dicono poco finché non si convertono in qualcosa di familiare. Come regola pratica, per l'italiano, **un token vale circa tre quarti di parola**.

Materiale	Circa
Una pagina di testo fitto (400 parole)	550 token
Un contratto di venti pagine	11.000 token
Una relazione di cento pagine	55.000 token
Un intero fascicolo da mille pagine	550.000 token
Il testo di un romanzo lungo	circa 200.000 token

Tradotto: una finestra da 200.000 token regge comodamente un fascicolo di trecento pagine più tutta la conversazione che ci fate sopra. Una finestra da un milione arriva a un archivio di progetto intero.

> **IL CONTO CHE SI DIMENTICA SEMPRE**
>
> Nella finestra non c'è solo quello che caricate: ci sono **anche tutte le domande e tutte le risposte** scambiate fino a quel momento.
>
> È il motivo per cui una conversazione lunga su un documento lungo si avvicina al limite molto prima di quanto suggerirebbe il conto delle pagine.

3.3 Lavorare bene su una conversazione lunga

Anche con una finestra ampia, una conversazione che va avanti per ore diventa faticosa da governare: il materiale si accumula, le richieste si contraddicono, e a un certo punto il risultato peggiora senza un motivo evidente. Tre abitudini risolvono quasi tutti questi casi.

1. **Una conversazione, un compito.** Quando si cambia argomento, si apre una conversazione nuova. Tenere insieme l'analisi di un contratto e la stesura di una newsletter non fa risparmiare tempo: fa peggiorare entrambe.
2. **Il riassunto di passaggio.** Prima di cambiare fase, si chiede: «riassumi in dieci righe quello che abbiamo stabilito finora». Quel riassunto si incolla come primo messaggio della conversazione successiva.
3. **Il materiale prima, la richiesta dopo.** I documenti si caricano all'inizio e non si cambiano più; le richieste si aggiungono in coda. È la disposizione che vedremo di nuovo nella sesta ora, parlando di riuso del contesto.

> **IL SEGNALE CHE È ORA DI RICOMINCIARE**
>
> Quando dovete ripetere una cosa che avevate già detto, la conversazione è diventata

troppo lunga.

Non è un difetto da correggere con la pazienza: è il momento giusto per aprirne una nuova con un buon riassunto.

Le domande che arrivano in aula

Sono le stesse in ogni corso. Vale la pena avere pronta una risposta breve e onesta, anche per quando le farà un collega.

«Qual è il modello migliore?»

Non esiste, e chi ve lo dice sta semplificando. Il più potente su un compito banale è spreco di tempo e di limite d'uso; il più leggero su un contratto è un rischio.

La domanda giusta è: **quanto costa sbagliare, qui?**

«Perché a volte si rifiuta di rispondere?»

Perché è addestrato su principi scritti, e una richiesta può assomigliare a un uso problematico anche quando non lo è.

Nella quasi totalità dei casi basta aggiungere il contesto: chi siete, perché vi serve, in che ambito professionale.

«Le risposte cambiano se rifaccio la stessa domanda?»

Sì, ed è normale: come abbiamo visto nella prima ora, la generazione è probabilistica.

Se vi serve un risultato costante, la strada non è ripetere la domanda finché esce giusta: è fissare il formato con i tag, che vediamo nella quinta ora.

«Conviene aspettare la versione nuova?»

No. Le versioni escono ogni pochi mesi e aspettare significa non partire mai. Quello che imparate — come si struttura una richiesta, come si sceglie un modello, come si verifica — vale su tutte le versioni.

3.4 Tre modi di dare il materiale, e quando usarli

Lo stesso documento si può fornire in tre modi diversi, con conseguenze pratiche diverse. Sceglierlo bene fa risparmiare molto tempo.

Il modo	Quando conviene	Da sapere
Incollare il testo	Poche pagine, materiale già in	Perde l'impaginazione e i numeri di

Il modo	Quando conviene	Da sapere
	formato testo	pagina: i riferimenti nella risposta diventano vaghi
Allegare il file	Documenti lunghi, PDF, tabelle, immagini	Mantiene struttura e riferimenti. È il modo predefinito per qualsiasi documento vero
Metterlo nella base documentale di un progetto	Materiale che serve in **tutte** le conversazioni su quel tema	Non va ricaricato ogni volta. È il tema della nona ora

L'ERRORE CHE COSTA PIÙ TEMPO

Ricaricare ogni mattina gli stessi cinque documenti di riferimento.

Se un materiale vi serve più di tre volte, il suo posto è la base documentale di un progetto, non l'allegato di una conversazione.

4. Esercitazione pratica dell'ora

Cinque scenari aziendali, un modello ciascuno. L'esercizio non è indovinare il nome: è saper spiegare la scelta.

Scenario	Modello scelto	Perché (una riga)
Classificare 5.000 email in arrivo per reparto	.....	.....
Scrivere la risposta a un reclamo importante	.....	.....
Riassumere il verbale di una riunione	.....	.....
Analizzare un contratto di fornitura	.....	.....
Estrarre i contatti da un elenco disordinato	.....	.....

IL CRITERIO DA SCRIVERE NELLA TERZA COLONNA

Non «perché è più bravo», ma: **che cosa succede se sbaglia?**

Se la risposta è «poco o niente», vai leggero. Se è «ci rimette un cliente» o «firmiamo una clausola sbagliata», vai potente.

Glossario

Le parole che incontri in questa dispensa, spiegate in una riga. Tienilo a portata di mano: torneranno in tutte le ore successive del corso.

Termine	Che cosa significa
Constitutional AI	Metodo di addestramento in cui il modello critica e riscrive le proprie risposte confrontandole con principi scritti.
Finestra di contesto	Quanto testo il modello tiene sotto gli occhi in una singola richiesta, documenti allegati compresi.
Haiku, Sonnet, Opus	I tre gradini della famiglia Claude: leggero, equilibrato, potente.
Public Benefit Corporation	Forma societaria che obbliga per statuto a perseguire anche una finalità di beneficio pubblico dichiarata.
RLHF	Apprendimento per rinforzo da feedback umano: persone valutano le risposte, il modello impara a produrne di più gradite.
Token	Il pezzetto di testo su cui lavora il modello: una parola, una sillaba, un segno di punteggiatura.

Per approfondire

Le fonti usate per costruire questa dispensa. Sono tutte pubbliche e consultabili: se un dato ti interessa, vai alla fonte e verificalo. È esattamente quello che chiediamo di fare anche all'intelligenza artificiale.

Il libro di riferimento del corso

Gianluca Bonomo, *Niente paura, è solo IA. La guida completa per capirla, usarla e restare umani*.

La documentazione ufficiale dello strumento

La pagina di supporto e la documentazione tecnica di Anthropic (support.claude.com e platform.claude.com). È l'unica fonte che resta allineata: nomi dei modelli, limiti e prezzi cambiano in mesi.

Dove verificare i modelli disponibili

Il menu di selezione dentro l'interfaccia di Claude mostra sempre le versioni attive per il tuo piano: è la fonte più aggiornata che hai.

UNA NOTA SULLE DATE

Le norme sull'intelligenza artificiale cambiano in fretta. I dati normativi di questa dispensa sono verificati a **settembre 2026**. Prima di usarli in una decisione aziendale, controlla sempre la fonte ufficiale: il calendario dell'AI Act è già stato modificato una volta.