

# Práctica 11

## Diseño ético y sostenible de un sistema de IA urbana (predicción tráfico urbano)

Jordi Blasco Lozano  
Computación de alto rendimiento  
Grado en Inteligencia Artificial

---

## Índice:

Índice:	2
1. Introducción y análisis del reto urbano	2
2. Tipos de datos necesarios y decisiones sobre su obtención	3
3. Elección y justificación del modelo de IA	4
4. Selección y justificación de la infraestructura HPC	4
5. Elección y justificación del modelo de IA	5
6. Reflexión ética y profesional sobre las decisiones tomadas	7
7. Comentario final	8

---

## 1. Introducción y análisis del reto urbano

Alicante, con más de 300 000 habitantes y gran afluencia turística, sufre congestiones frecuentes en horas punta (autovía A-70, sobre todo en la salida de Mercalicante desde Elche y también por las avenidas urbanas) que reducen la eficiencia del transporte y aumentan la contaminación. La Inteligencia Artificial combinada con Computación de Alto Rendimiento puede anticipar los picos de tráfico y optimizar semáforos o rutas alternativas en tiempo real, mejorando la movilidad y reduciendo emisiones. Por ejemplo, simulaciones basadas en HPC han demostrado que se pueden conseguir menos atascos y tiempos de viaje reducidos gracias a optimizar la red de transporte. No obstante, este enfoque introduce dilemas éticos iniciales: la recolección masiva de datos ciudadanos plantea riesgos de privacidad, y las soluciones deben ser equitativas, evitando favorecer sólo a determinados barrios o grupos sociales. Así, el proyecto debe equilibrar el beneficio en movilidad con criterios de sostenibilidad, protección de datos y justicia urbana.

## 2. Tipos de datos necesarios y decisiones sobre su obtención

Para modelar el tráfico se requiere integrar datos heterogéneos en tiempo real o cuasi-real. Un sistema de ciudad inteligente puede ingerir millones de eventos por minuto (sensores, móviles, clima, etc.), por lo que se necesitan fuentes robustas y protocolos de privacidad. Entre las principales fuentes se incluyen:

- **Sensores de tráfico fijos:** bucles inductivos en carreteras y avenidas, detectores de paso en semáforos y cámaras de vigilancia. Estos dispositivos proporcionan flujos de vehículos, velocidades medias y ocupación de carriles. Son controlados por la DGT y el Ayuntamiento. Se deben anonimizar los datos, contar vehículos y encriptar la transmisión para poder procesar estos datos.



- **Datos móviles agregados:** información de localización anónima de teléfonos y aplicaciones de navegación (Google Maps, Waze, móviles de la ciudadanía). Estos datos permiten inferir trayectos y densidades de tráfico por zonas. Se obtendrían a través de acuerdos con operadores o empresas de smart mobility. Por su carácter sensible, es imprescindible usar sólo datos agregados y anonimizar patrones de movimiento para cumplir con la ley de protección de datos.
- **Eventos y características urbanas:** calendario de eventos masivos (conciertos, congresos, partidos) y obras públicas que generan atascos puntuales. Además, datos meteorológicos (precipitación, viento, temperatura) influyen en el tráfico. Estas fuentes se obtienen de organismos públicos (AEMET, agenda municipal) o sensores IoT urbanos. Se incluirán en el modelo con etiquetado de eventos sin asociar identidades.
- **Otros datos de referencia:** informes históricos de congestión urbana, mapas de carreteras y semáforos, datos socioeconómicos (zona industrial, colegios) para contextualizar el modelo.

Cada tipo de dato se selecciona sopesando calidad y privacidad. Por ejemplo, se prioriza usar números de flujo de personas en vez de información individual de pasajeros. Además, se implementarán medidas de mitigación: agregación estadística de los datos (eliminando ID personales), protocolos de cifrado extremo a extremo, y cumplimiento estricto de la ley de protección de datos. De este modo, se garantiza que la IA trabaje con información suficiente para predecir congestiones, pero sin vulnerar la privacidad ciudadana.

---

### 3. Elección y justificación del modelo de IA

Dada la naturaleza secuencial y multivariable del tráfico, se opta por un modelo avanzado de red neuronal recurrente LSTM (Long Short-Term Memory), capaz de captar dependencias temporales complejas. Como alternativa simple se compara con ARIMA, un modelo estadístico clásico. ARIMA es adecuado para series con estacionalidad marcada (por ejemplo, consumo horario diario). Funciona con datos estables, requiere poco entrenamiento y es altamente explicable ya que usa fórmulas lineales claras.

Sin embargo, ARIMA solo modela relaciones lineales en una variable y no maneja bien múltiples factores externos (clima, eventos). Por el contrario, LSTM puede procesar secuencias con varias entradas simultáneas y aprender patrones no lineales a largo plazo. Técnicamente, ARIMA no necesita GPU y consume muy poca energía, lo que lo hace eficiente y transparente. En cambio, LSTM exige potencia de cómputo mayor haciendo uso de un entrenamiento intensivo y sus decisiones son una “caja negra” difícil de explicar. Éticamente, esta opacidad se contrarresta con métodos de interpretabilidad (por ejemplo, análisis de sensibilidad o LIME) para validar los resultados. Se justifica la elección de LSTM porque la predicción de tráfico en Alicante se ve afectada por múltiples variables externas y patrones irregulares. La ganancia en precisión y capacidad predictiva de LSTM supera la mayor complejidad, dado que un modelo más simple podría pasar por alto escenarios críticos. No obstante, se reconoce que ARIMA serviría como línea base ligera para contextos donde la capacidad de cómputo o la sostenibilidad energética sean prioritarias. En resumen, LSTM se elige por su precisión y flexibilidad, mientras que ARIMA se considera como alternativa más ligera en términos de consumo y transparencia.

---

### 4. Selección y justificación de la infraestructura HPC

El modelo LSTM requiere recursos de cómputo intensivos. Por ello se propondrá usar infraestructura HPC con GPUs. Concretamente, un clúster de entrenamiento podría consistir en varios nodos equipados con GPU (p.ej. NVIDIA A100) y CPUs multinúcleo, interconectados por Red de alta velocidad. Alternativamente, se pueden usar servicios de HPC en la nube (AWS, Azure ML) o incluso supercomputadores nacionales (por ejemplo, MareNostrum en Barcelona). La HPC distribuida permitirá paralelizar el entrenamiento de redes y procesar flujos de datos en tiempo real (detección de accidentes, streaming de sensores).

El cómputo requerido se estima del orden de decenas de TFLOPS. Por ejemplo, entrenar un LSTM con varios años de datos para 10 distritos podría llevar alrededor de 4 horas por iteración en una GPU de 300W. Con 4 GPUs trabajando en paralelo (1.2 kWh por GPU por sesión), cada iteración consumiría 4.8 kWh. Usando 10 iteraciones, el gasto total rondaría las decenas de kWh. Este cálculo justifica el uso de una infraestructura distribuida: si se usara un único PC normal, el entrenamiento tardaría mucho más.

Desde el punto de vista energético, se adoptarán criterios de eficiencia: uso de GPUs modernas con alto rendimiento por vatio, posibles horarios nocturnos de entrenamiento cuando la red eléctrica tenga más energía renovable, y ajuste de precisión (entrenamiento en FP16) para ahorrar energía. Además, se prevé implementar paralelismo optimizado (OpenMP/MPI o CUDA) para escalar en múltiples núcleos. Se valorará el cómputo en la nube HPC como opción escalable, aunque con precauciones: la nube implica costes económicos y cuestionamientos de soberanía tecnológica (p.ej. ubicación de datos). Si se dispone de un clúster local (por ejemplo en la Universidad de Alicante o en colaboración con el BSC), se minimizaría la latencia y se podría optimizar el consumo siguiendo estándares HPC de eficiencia energética.

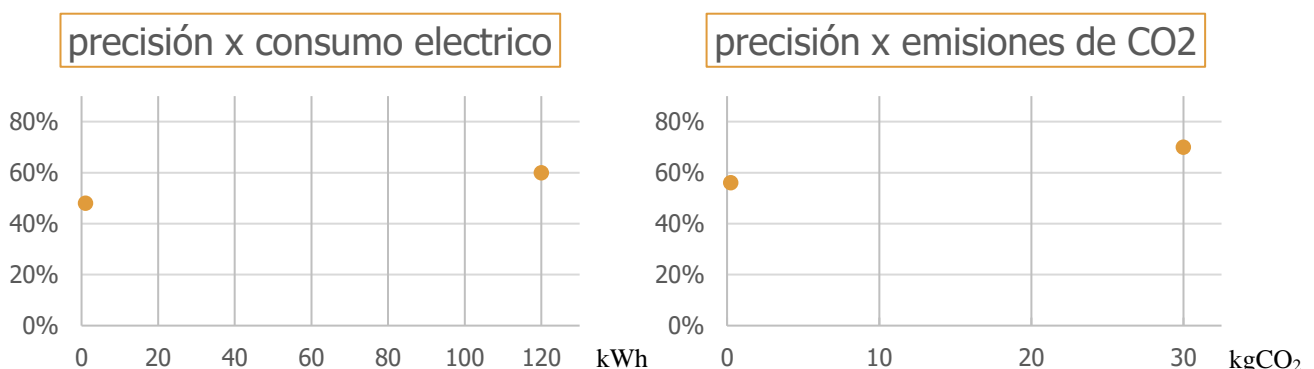
## 5. Elección y justificación del modelo de IA

A modo orientativo, estimamos el consumo de entrenar el modelo LSTM y lo comparamos con ARIMA. Suponiendo un entrenamiento de LSTM de 4 h en una GPU de 300 W (1.2 kWh/sesión) repetido 10 veces para ajustar parámetros y en 10 zonas, el total sería:  $4h \times 300W \times 10\text{sesiones} \times 10\text{zonas} \approx 120 \text{ kWh}$ . Para visualizar, 120 kWh equivale aproximadamente a cargar un coche eléctrico de unos 600 km de autonomía o a 2 semanas de luz en un hogar promedio.

En contraste, el entrenamiento de un ARIMA (en un PC de 100 W durante 1 h por sesión) resultaría en  $0.1 \text{ kWh/sesión} \times 10\text{sesiones} \times 10\text{zonas} = 1 \text{ kWh}$ . Por tanto, la red LSTM consume unas 120 veces más energía que el modelo estadístico.

En términos de emisiones de CO<sub>2</sub>, si el mix eléctrico de Alicante es de aproximadamente 250 gCO<sub>2</sub>/kWh, el entrenamiento completo de LSTM genera  $\approx 30 \text{ kg CO}_2$ , mientras que ARIMA solo 0.25 kg. Aunque estas emisiones no son enormes en sí, un despliegue a gran escala (entrenamientos frecuentes, múltiples simulaciones de escenarios) multiplicaría el impacto. Por ello, se justificaría usar el modelo LSTM solo cuando la precisión adicional lo demande; en casos con recursos restringidos se privilegiaría el modelo sencillo para ahorrar energía.

Para visualizar de una mejor forma los datos anteriores he creado dos diagramas de dispersión en los cuales podemos observar el costo de energía y emisiones correlacionado con la precisión de cada modelo.



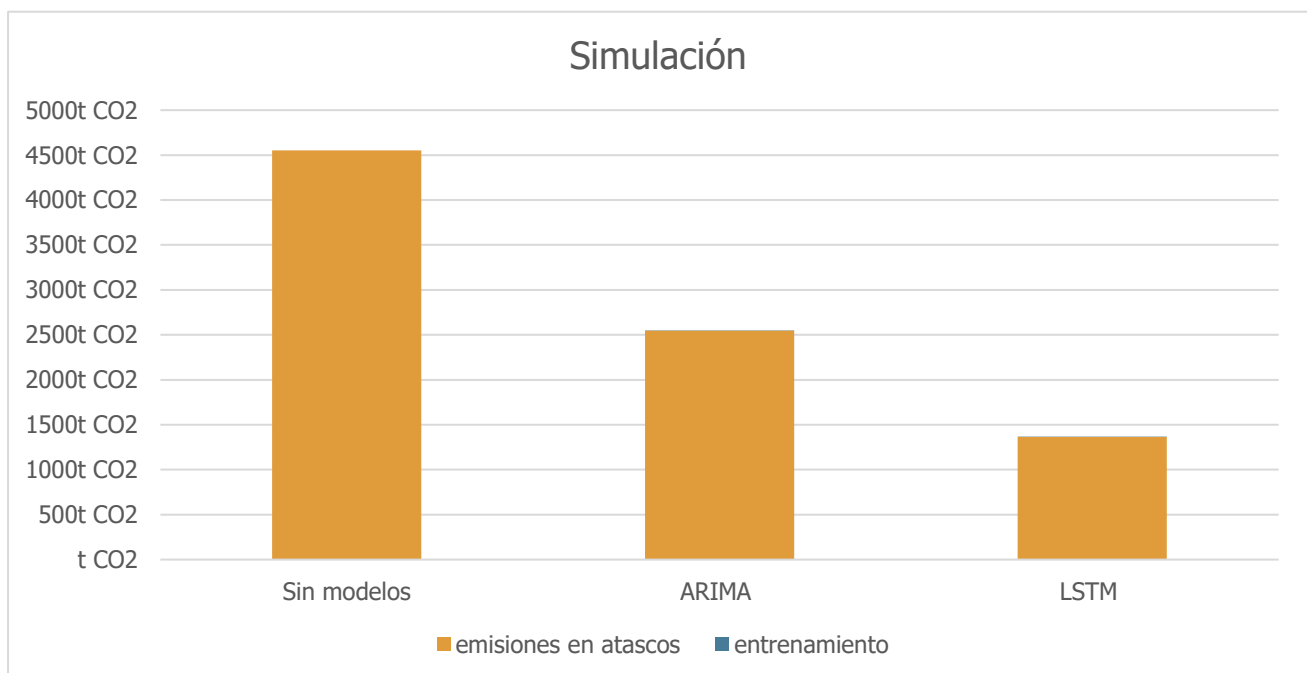
En estas tablas he usado un dato ficticio como base para sacar las precisiones, he tomado el modelo LSTM y le he otorgado un 70% de precisión en disolución de atascos, para el modelo ARIMA le damos un 20% menos de precisión sobre estos 70%, es decir un  $70 \times 0.8 = 56\%$ .

Para mejorar la sostenibilidad, se explorarán técnicas adicionales: early stopping para reducir iteraciones inútiles, podar la red o aplicar cuantización después de entrenar, y reutilizar modelos pre-entrenados con aprendizaje continuo en lugar de reentrenar desde cero. Estos enfoques, junto con la selección de hardware eficiente, buscan minimizar el coste energético manteniendo la calidad del servicio.

Para hacer el problema aún más realista podríamos hacer una estimación de las emisiones de CO<sub>2</sub> que se producirían con el 100% atascos y comparar las emisiones que se generarían con los modelos en funcionamiento (sumando también las emisiones de entrenamiento y procesamiento de cada modelo)

Para hacer esto haremos uso de la IA para que nos proporcione una estimación del número de emisiones mensuales de más que generan los atascos en alicante. Este número será la resta de emisiones totales con atascos menos las emisiones totales con atascos.

La respuesta obtenida ha sido de 4 551 toneladas de CO<sub>2</sub> al mes, vamos a suponer que entrenamos el modelo de LSTM y el de ARIMA unas 10 veces por mes. Esto nos da 300kg de CO<sub>2</sub> para LSTM y 7,5kg para ARIMA. Con las efectividades dichas anteriormente del 70% y del 56% respectivamente, aun con los modelos funcionando habría atascos por lo que el modelo LSTM a una efectividad del 70% la ciudad seguiría emitiendo 1365,3 toneladas de CO<sub>2</sub>, mientras que el modelo ARIMA 2548,56 toneladas de CO<sub>2</sub>. Con estas estadísticas haremos una gráfica que expresen de una mejor forma todos los resultados.



Como observamos en esta simulación, las emisiones del entrenamiento de los modelos comparado con la cantidad de toneladas de CO<sub>2</sub> que producen los atascos es casi imprescindible. Con esto debemos concluir dos cosas, la primera sería que si conseguimos que estos modelos tengan una precisión (aunque sea poca) sería de una gran ayuda para la mejora de la calidad del aire de la ciudad. Lo segundo que podemos concluir es que debemos de tener en cuenta más variables a este problema tales como el dinero que costaría la implementación, ya que el consumo de los modelos no constituye un problema de emisiones, más bien todo lo contrario.

## 6. Reflexión ética y profesional sobre las decisiones tomadas

- **Equidad social:** El sistema debe beneficiar a toda la ciudad, sin concentrar mejoras solo en zonas céntricas. Se vigilará que el modelo no ignore barrios periféricos con menos sensores o menor tráfico histórico. La planificación propuesta debe considerar el acceso igualitario: por ejemplo, las optimizaciones de semáforos no deben perjudicar rutas de transporte público ni forzar a conductores a rutas peligrosas. En general, la IA debe respetar criterios de justicia urbana, asegurando que los beneficios (la reducción de atascos) se distribuyan de manera razonable entre todos los ciudadanos.
- **Privacidad y transparencia:** Al recopilar datos de vehículos y móviles, se debe de cumplir la normativa de protección de datos. Toda información es anónima y agregada, sin ningún enlace a identidad personal. Se debe de documenta el uso de datos y se informará a la población sobre el propósito del sistema. Además, se mantendrá la transparencia en los criterios de decisión. Aunque LSTM sea una “caja negra”, se podrían implementar herramientas de explicabilidad (p.ej. mapas de calor de variables importantes) para que los gestores entiendan por qué el modelo predice congestiones. Esta transparencia reforzaría la confianza social y el control ético del sistema.
- **Explicabilidad y responsabilidad profesional:** Los diseñadores del sistema deben ser capaces de justificar técnicamente las decisiones (elección de LSTM, ajustes de parámetros) y sus implicaciones. Se deberían de incluir los análisis comparativos para mostrar el balance entre precisión y sostenibilidad. Cualquier recomendación de cambio en la infraestructura de transporte (p.ej. nuevos semáforos o desvíos) será validada mediante simulación HPC previa, siguiendo un enfoque basado en evidencia. Los desarrolladores y responsables políticos deben de ser conscientes de su responsabilidad de evitar resultados perjudiciales, tales como accidentes, desigualdades, etc.
- **Sostenibilidad ambiental:** Aunque la HPC consume mucha energía el sistema pretende reducir emisiones globales al evitar congestiones. Se realizará un análisis comparativo **real** del beneficio neto: la disminución de CO2 vehicular gracias a flujos optimizados vs. la energía empleada en cómputo. Se fomentará el uso de fuentes renovables para alimentar los centros de datos cuando sea posible. En cualquier caso, se busca un diseño sostenible: si la mejora en movilidad es marginal, se privilegiará la opción más eficiente (por ejemplo, usar ARIMA o reducir la frecuencia de reentrenamiento) para minimizar la huella ecológica del sistema (si la simulación anterior no fuera correcta).
- **Formación y talento:** Finalmente, se reconoce que este proyecto requiere personal cualificado en IA y HPC. Se promoverá la capacitación técnica del equipo (posibles colaboraciones con universidades o centros HPC). La escasez de talento en HPC es un reto actual, por lo que el proyecto servirá también para fomentar el conocimiento local en computación paralela y eficiente.

En resumen, las decisiones técnicas se guían no solo por la factibilidad, sino por un compromiso ético: proteger la privacidad, favorecer la equidad y minimizar el impacto ambiental, mientras se busca eficiencia en el diseño profesional.

---

## 7. Comentario final

El diseño presentado es adaptable a otras ciudades o contextos similares. Ciudades medianas en España con retos de movilidad (por ejemplo, Valencia, Murcia o Vigo) podrían aplicar este sistema cambiando únicamente las fuentes de datos locales. Además, el enfoque es transferible a otras problemáticas urbanas: por ejemplo, un gemelo digital similar podría prever la ocupación de zonas verdes o el consumo eléctrico por barrio, utilizando datos análogos (sensorización IoT, históricos) y ajustando las variables de entrada. Iniciativas globales avalan esta adaptabilidad: Virtual Singapore, por ejemplo, creó un gemelo digital con tráfico, clima y planos BIM para simular efectos de nuevas infraestructuras. De modo análogo, Alicante podría ampliar el modelo para simular la apertura de una nueva vía o el cierre de calles al tráfico. Por último, las metodologías de IA+HPC propuestas (modelos predictivos, análisis masivo) son escalables a proyectos urbanos de mayor envergadura, siempre que se dimensione apropiadamente la infraestructura de cómputo. En definitiva, este sistema sirve de plantilla replicable para cualquier ciudad que busque soluciones inteligentes y sostenibles a problemas de movilidad urbana.